

Appraising Information in Machine Language Form

By MEYER H. FISHBEIN

DURING A RECENT conversation about the state of the computer art, one expert expressed his opinion that the computer revolution had ended. Now all institutions must adapt themselves to the multivariate spinoffs that the revolution generated, while computer capability will continue to expand. To reject this judgment we would have to assume a revolution with no foreseeable end. To the extent that archival organizations are concerned with current record practices and the content of future holdings, they cannot long delay plans for appraising, accessioning, and servicing this new media of communication.

Archivists must meet the challenge to deal with this problem, though few feel fully qualified. This is borne out in conversations with both foreign and domestic colleagues. Despite my occasional attention to developments in electronic devices for about 5 years and fairly intensive study for the next 5, I disclaim complete confidence in my capacity to make judgments that will stand the test of time. Nevertheless, archivists must apply professional talents to appraise, and thus preserve, machine-readable records "with all deliberate speed." Unless we fulfill our responsibilities, about one million reels of tape in the Federal Government and more elsewhere will be erased without any archival judgments on the continuing value of the information they store.

It was about 10 years ago that someone in the National Archives and Records Service asked my opinion about the appraisal of information on electronic tape. Without hesitation I replied that tape containing data that originated in a Government agency was a record, subject to the same appraisal standards as any other record. Whether tape was a proper medium for indefinite storage was of secondary,

The author, a Fellow of the Society, is Director of the Records Appraisal Division, National Archives. He read this paper on September 30, 1970, at the SAA 34th annual meeting in Washington, D.C.

technical interest. In any case, we could insist on a complete print-out as a last resort.¹

By 1964 most Federal agencies had computerized systems, and many humanists and social scientists joined business managers and physical scientists in revolutionizing their collection, processing, and dissemination of information. At this point I devised my second maxim on the subject: "Unless archivists react positively to the computer revolution archival establishments would be limited to the records of the *ancien régime*."

Thus, I had offered two seemingly disparate views: the machine language media should be treated like any other documentary material; and the medium will have a revolutionary effect on archives. Both propositions are reasonably valid for an analysis of where we stand on the appraisal of the media.

My first proposition refers to punched cards, tape, discs, drums, and the like, as a record type, although many archivists continue to question the record character of information in machine language form. This medium, they argue, merely transcribes and processes data from one traditional form, the source documents, to another traditional form, a publication or report. They cite the precedent of several decades during which punched cards were designated as "non-record."

The earliest reference I found to this precedent was a memorandum by Arthur H. Leavitt of July 29, 1936, that summarized an interview with the Chief Clerk of the Census Bureau.² The Chief Clerk asked Dr. Leavitt whether the National Archives had any interest in the approximately 8 million census and survey cards that had been tabulated. Both agreed that all conceivable use of the cards had been exhausted. In Leavitt's opinion the cards were "nonrecord." Archival values for the data were satisfied by the final tabulations and, in some cases, the original questionnaires. Though we may well speculate what our policy would have been had Leavitt discovered valuable untabulated data transcribed from questionnaires that had been destroyed, this and other similar decisions became policy in regard to punched cards.

In 1960 the Federal Records Council suggested that NARS ex-

¹ On occasion this method has been followed. The tabulations of foreign commerce statistics for the 1930's that were prepared from punched cards for the annual *Foreign Commerce and Navigation of the United States* were accessioned as a complete printout. The value of miniaturization is lost when this method is followed.

² Memo No. 87, Chief, Division of Department Archives, to Director of Archival Services, July 29, 1936, Central Files, 1935-44, Records of the National Archives and Records Service, RG 64, National Archives. For a further discussion of the development of appraisal standards see Meyer Fishbein, "Appraisal of Twentieth Century Records for Historical Use," in *Illinois Libraries*, 52:154-162 (Feb. 1970).

amine the problems relating to the disposition of data in machine language form. The preliminary study, prepared by Richard Jacobs, now with the Office of Presidential Libraries, recommended that we consider electronic tape as "interim media" and that essential information be retained in conventional form. It is noteworthy that Mr. Jacobs cited Census Bureau tape files, which, "like the punch cards for earlier censuses, will lose their usefulness in time."³ Jacobs's 1961 study was prepared before the development of source data automation, data banks, and machine language information systems. By 1965 our attitude toward these media had changed radically.

Many researchers were then desirous of actually converting historical quantitative information from conventional to machine language form. Erasing data of continuing utility already on tape would have been a disservice to researchers. Furthermore, some of the taped files were the basic communication medium for carrying out the functions and activities of agencies, and as such were official records. Fortunately, Federal legislation did not require amendment to include tape in the definition of records. The Federal Disposal Act of 1943 defines records as "books, papers, maps, photographs, or other documentary materials regardless of physical form or characteristics, made or received by an agency"⁴ Yet, some State legislation may require amendment because it exclusively defines conventional materials, usually books, papers, and maps. Ohio amended its laws in 1959 to define records as "any document, device, or item created or received by . . . a public office . . . to document the organization, policies, or other activities of the office."⁵ No one can reasonably deny that the final tapes being generated by the 1970 census are records within the meaning of Federal and Ohio laws.⁶

Like other communication media, machine language devices may also be designated "nonrecord" and therefore not statutorily subject to archival appraisal. They are often equivalents of conventional reference materials and workpapers. With regard to the latter, scratch, testing, and program tapes and media used exclusively for processing data through a system are nonrecord materials with a short life expectancy. Disposition decisions concerning these materials should be left to the data processing staff.

Tapes and punched card decks that contain scientific, economic, and social data are received by many types of institutions—govern-

³ "Training Lectures No. 56, On Records Preparation and Magnetic Tape" (unpublished, National Archives and Records Service, 1960), p. 1, 5.

⁴ 57 Stat. 380.

⁵ Ohio Revised Code (1959), § 149.40

⁶ See also Oliver W. Holmes, "Public Records"—Who Knows What They Are?" in *American Archivist*, 23:5-26 (Jan. 1960).

mental agencies, universities, and businesses—as reference sources for analysis and direct use. The County Data Tape Files of the Census Bureau, for example, are sold to numerous organizations for inclusion in a data bank or reference library. These tape files are no more a record of the user than a conventional publication. But, under certain circumstances, processing tape and machine language publications have record status with potential archival value.

Processing data through several runs may be designed not only to correct and merge data but to purge information that is not to be included in final reports. When the purged data have enduring value, the intermediate tape with all the data should be selected for preservation. Similarly, unique and valuable information may be added to tape purchased from another organization, thus creating a record subject to archival appraisal. A conventional census publication that is annotated by an official at the policymaking level to note that the statistics show the necessity for massive aid to Appalachia or corrected to show data received directly by the purchasing agency should be treated as a record. 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.

To this point, my early reference to the *ancien régime* appears irrelevant. “*Plus ça change . . .*” seems to be the appropriate French expression. Yet imminent changes in the content of archival establishments are probably more significant than what remains the same. Because an enumeration and explanation of all changes already evident could occupy considerable time, I am limiting myself to a few significant developments.

The potential market for archival sources must determine our appraisal judgments. With the extensive use of the third and introduction of the fourth generation of computers, demands for source data are increasing exponentially. That is, each study and policy decision based on computer techniques generates several other studies and decisions, which result in the creation and dissemination of additional source data files that encourage even more computerized studies. We may then project a substantial increase in the use of archives to the degree that we provide for the preservation of useful data sources and analysis of these sources.

During the decade of experimentation on the first generation of computers, to about 1955, few archivists could foresee future holdings of machine language media. The earliest applications, the formulation of firing tables and other rapid mathematical calculations, seemed to have no archival potentials. My own chief interest in the role of

the computer at that time was directed toward the preservation of records on computer technology.

Retrospectively, however, we may detect portents of some present applications. The firing tables were designed to make decisions at the lowest military level. Secretary of Defense Robert McNamara used more sophisticated machines for decisionmaking by rapid calculation of benefits from various alternative allocations of men and materiel. Final judgment of his methods must await evaluations of his data input, the assumptions underlying his formulation of Defense problems, and the results achieved. Computer related records—the original and final data tapes, feasibility studies, and basic program instructions—should document these decisions. Computer use for decisionmaking has obvious implications for any recordkeeping system. Rapid access to information has already increased reliance by executives on data sources as contrasted with their predecessors who usually ignored similar data in relatively inaccessible conventional form. When important decisions are based on computerized data, the sources may be accessionable. In the past, Archives rarely located the sources for policy decisions.

The second generation of computers permitted a wider range of calculations, with some memory storage and capability to perform several operations simultaneously. The Census Bureau had experimented with computers for aggregating voluminous source data. By 1955 its highly skilled staff had developed techniques for complex aggregations and correlations with increased speed and accuracy.

Econometricians soon recognized the utility of the media for model-building and other economic analyses. One economist estimated 3 years ago that about half of all economic studies had been prepared during the previous decade.⁷ Their demands created new data sources in the Federal Government.

As econometric studies are useful for economic planning and prognostication, demands are being placed on such agencies as the Census Bureau, the Office of Business Economics, the Economic Development Administration, the Federal Reserve System, and the Bureau of Labor Statistics. Each of these agencies has organized large tape libraries with holdings that require appraisal.

Economic historians followed the econometricians in organizing vast data sources by computer techniques for analysis of historical questions, such as the profitability of slavery, the roles of railroads and capital formulation in American economic development, and the elements of industrialization during the 19th century. Within a few years, during the late 1950's and early 60's, the traditional relativists

⁷ *Business Week*, Jan. 6, 1968, p. 7.

among economic historians, who rarely quantified their judgments and infrequently sought source documents, were on the defensive. Political and social scientists and historians are also relying increasingly on empiric evidence to test and develop theories and generalizations. This trend, if it continues, will result in greater reliance on archival holdings.

These new demands for historical data raise important questions about the accuracy, accessibility, and relevance of current data. Sociologists, urbanologists, and many Federal executives are reviewing their needs for source data and bench marks in the fields of education, housing, employment, income, vital statistics, health, progress in racial and ethnic adjustments, delinquency, consumer practices, and other indicators of our social condition. Social indicators, they claim, are more urgently required than economic indicators. These demands will be met by the collection, compilation, and analysis of numerous data sets.

Archivists must be prepared to deal with the great increase in information sources. Appraisal problems as regards machine language records will, for a time, be principally concerned with the source data files that are enumerated above. Source data will, to some extent, be supplied by the administrative process. It is this class of statistics that has been the subject of investigations because improper release may invade privacy. While recognizing his obligation to protect privacy, the archivist may look forward to the time when such source data can be used for studies without harming individuals.⁸

In appraising statistical sources, archivists may adopt the following procedures:

1. Where possible, appraise all files relating to each survey before it is completed. Delay may result in loss of valuable information by erasure of the tape in order to reuse it.⁹
2. Locate essential methodological and program documentation to assure its preservation. If either is missing, all related files may be valueless.
3. Compare the data in the source documents with the raw data tape. When all data of long term interest are transcribed to tape, the source data records (usually questionnaires) should normally be scheduled for disposal.
4. Compare content of corrected raw data tape with that on intermediate tapes to determine the value of purged information. In certain cases, two or more files relating to a single survey may be retained for different research purposes.

⁸ For an interesting discussion of this point see Arthur R. Miller, *The Assault on Privacy: Computers, Data Banks, and Dossiers* (Ann Arbor, 1971).

⁹ Erased tape has a value of approximately one half its original price. Unlike paper, which has little value once used, tape presents a problem to the archivist, who may find himself competing with a cost-conscious administrator wanting to reuse the tape.

5. Eliminate unnecessary duplication of data. Frequent exchange of information among agencies and within an agency creates redundant sources of temporary value.

6. Decide whether both the final survey report and the final tabulations in machine language form should be retained. They are, in fact, two distinct records.

7. When the appraisal proves difficult, consult with data banks and individual experts who specialize in the subject area. In the case of random or stratified samples of a population, also consult a qualified statistician.

8. Determine whether files of merged data from two or more sources have sufficient research interest to justify retention.

9. When tape is updated by replacing earlier data by current data and the information has value for studies of long term trends, consider the advisability of freezing the information periodically.

The earliest standard for selecting permanently valuable machine language records appears in the retention plan for the Census Bureau.¹⁰ Although this provision was prepared in 1965 before archivists had developed any clear standards, the plan is still being reasonably administered. In brief, it provides for retention of computer plans, feasibility studies, and final master tapes and/or summary punched cards containing demographic, social, political, or economic data that have enduring value. A recent schedule by the Bureau, concerning business censuses, provides for the preservation of final corrected master data tapes and "procedural records essential for using the master data tapes."¹¹

The standards for preserving statistical sources may be applied to other computer media. We have agreed to accept, for example, the biochemical and psychological data concerning the scientists who inhabited a capsule in the sea near the Virgin Islands. Although the information is being studied intensively by experts, it does not appear to us that the detailed recordings of all behavioral patterns will be fully analyzed. New interpretations of the original data will, we believe, be produced from time to time. Basic NASA recordings from satellites also probably have archival value. The problems of appraising raw scientific data are very complex and highly specialized. NARS is seeking the aid of the National Academy of Sciences in obtaining advisory panels.¹²

Certain appraisals will be based on an informed belief that tape will, in most cases, replace traditional hard copy. We can foresee

¹⁰ Department of Commerce, "Retention Plan for Records of the Bureau of the Census," 1965, sec. 30.

¹¹ Census Bureau schedule for Disposal Job No. NN-170-143, NA.

¹² See Theodore McAllister, "Collecting Archives for the History of Science," in *American Archivist*, 32:327-332 (Oct. 1969).

that land title searches will be computerized; therefore, the taped recordings of land transactions should be designated for preservation.

An unfortunate assumption exists among academicians, managers of information systems, and generators of quantitative data that archivists are unconcerned about the preservation of machine language records. Recent events show otherwise. In 1969, the National Archives and Records Service established a Data Archives staff in the Office of Records Management to survey Federal agency holdings and locate accessionable tape files. Shortly thereafter the staff accessioned several tape files, and it became a branch within the Office of the National Archives. The SAA Committee on Data Archives and Machine-Readable Records plans to canvass other domestic archives to determine what plans are being developed for the preservation of data files.

Foreign archivists are also developing plans for accessioning machine language records. This activity was considered at the International Archives Congress of the International Council on Archives in 1964 and the London ICA Round Table in 1965. In August 1971, the Council distributed a detailed questionnaire about the state of the art because written tradition "is on the brink of a complete mutation and . . . the consequences of this phenomenon are vital to archives."¹³

Let me conclude with a personal view of where we seem to be going. Data sources of enduring value for all agencies will be centralized in archival establishments. Within the Federal Government, unless Congress decides otherwise, NARS will provide this centralized storage, thus increasing the value of machine-readable data from any one agency by merging and correlating them with information accessioned from other agencies.

Archives will accession finding aids in machine-readable form although they are not records. Index cards were often denied to archives because they could not be located or they were not identified. Tape, on the other hand, is labeled and controlled by libraries.

Archives will communicate and transfer their information to researchers by remote control circuits. The use of the computer will facilitate and possibly democratize research. Students will no longer have to spend long periods in archives. Universities will have access to hardware that will be available to students and faculty.

The establishment of specialized data banks and information centers will provide support for the archival institutions that acquire source data.

¹³ Questionnaire on "Archives and Automation" issued in preparation for the 13th International Round Table on Archives at Bad Godesberg, Federal Republic of Germany, September 1971.

Archives will be used increasingly for current planning by analyzing computerized social, economic and administrative data. Analysis is considerably less costly than data gathering.

Although retention plans for agency tape are useful in alerting the agency about the need to select files for preservation, important appraisals will continue to require *ad hoc* decisions after intensive study of each computer system.

In 1967, David Easton, Chairman of the Committee on Information in the Behavioral Sciences of the National Academy of Sciences, wrote that "the whole communication process is undergoing a radical transformation, and this inevitably brings with it an increasing volume of information."¹⁴ Archivists, working with the necessary cooperation of records officers and ADP technicians, must prepare themselves for this radical transformation or restrict themselves to certain conventional files. Even these files may not be completely comprehensible without reference to related computer sources.

¹⁴ National Academy of Sciences, *Communications Systems and Resources in the Behavioral Sciences*, p. v (New York, The Academy, 1967).

Archival Enrichment

In 1970 the State Records Section of the Division of Archives and Records Management lent its paper shredding machine to North Carolina State University for use in an experiment. Conducted jointly by the departments of Civil Engineering and Soil Science, the experiment was designed to determine if shredded paper and other carbon-producing waste matter could be mixed with nitrogen-producing poultry waste to create an economical soil enricher. The professors and graduate students participating in the project developed a mixture which, when applied to plants, increased their dry weight by ten times and actual food production by four times. The mixture can be produced cheaply, for about \$15.00 a ton.

It is hoped that a new outlet has been found for the great quantities of wastepaper generated by virtually all countries of the world and that a cheap soil builder and fertilizer has been discovered for use in underdeveloped countries.

—from *Carolina Comments* (North Carolina State Department of Archives and History), 19:106 (Nov. 1971).