

January 1990

Adding Electronic Records to the Archival Menagerie: Appraisal Concerns and Cautions

Michael E. Holland
Oregon State University

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/provenance>



Part of the [Archival Science Commons](#)

Recommended Citation

Holland, Michael E., "Adding Electronic Records to the Archival Menagerie: Appraisal Concerns and Cautions," *Provenance, Journal of the Society of Georgia Archivists* 8 no. 1 (1990).

Available at: <https://digitalcommons.kennesaw.edu/provenance/vol8/iss1/4>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Provenance, Journal of the Society of Georgia Archivists by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

Adding Electronic Records to the Archival Menagerie: Appraisal Concerns and Cautions

Michael E. Holland

Electronic records are significantly different from most of the records held in institutional archives and, thus, they must be appraised and evaluated for accessioning with different and additional factors in mind.¹ The archival predilection for discussing electronic records only within

¹ Electronic records for the purposes of this paper are defined as records that are stored or maintained in a machine-readable form and require the intervening use of a computer to render the information in human readable form. Electronic data is encoded in binary code and includes, but is not limited to, the following physical forms: magnetic tapes, magnetic tape cartridges, disk packs, floppy diskettes, magnetic cards, hard disk drives, and optical disks.

the context of preservation issues fosters the misconception that—given optimum environmental conditions—electronic media may be considered an archival or long-term storage format. This is not a realistic assessment. Electronic media should be approached as a transitory information format in the archives. The mistaken beliefs that electronic recording formats are long-term storage media and that the readily apparent and sometimes superficial advantages of the volume-to-data ratio of electronic records when compared to other formats have led to the excessive accessioning of machine-readable records into some institutional archives with inadequate reasons to justify the transfer.

It is difficult to discuss the preservation of machine-readable records, because the professional literature is replete with warning and cautionary statements about physical control and care, but spare in discussing the decision-making process that will result in the presence or absence of electronic records in the archival repository. It is the appraisal process that must be given more and earlier attention, if archivists are to add the proper and most valuable electronic records to their holdings and devote to them the proper amount of resources and care. If marginally valuable records are accessioned and preserved at great cost to the archival program, the result is the denial of resources and care to other records already within the archives or the inability to accept worthy records into the repository because of inadequate resources. Or if historically valuable records are accessioned into the archives in machine-readable form, but over time become inaccessible to the user, the result is

the same as if the records had been disposed of prior to being brought into custody.

Electronic record media—upon or prior to removal to the archives—requires either immediate transfer of information to more physically stable and long-lived media or periodic and scheduled recopying onto fresher stock of the same medium. Accessioning electronic media into an institutional archives requires either an investment of resources at the time of accessioning in order to transfer records stored on transitory media to more stable media or the continuing investment of resources to maintain records in electronic or machine-readable format. It is thus critical that an archivist contemplating the accessioning of electronic records consider a minimum of five fundamental factors. These factors will be dealt with in more detail later in the article and will be posed in the form of interrogatives to make them easier for the archivist to apply to individual record series.

In discussing physical and technical factors that will affect the appraisal of electronic records for archival disposition, it is essential to consider briefly the physical forms of the records. Tape is still the most heavily used storage medium for large data files. Its primary use is in database applications, and it is used as both an input and output source for mainframe computers. Half-inch computer tape comes in a variety of data density and tracking formats: 7, 9, and 11 track tapes, and 800, 1600, 6250, and 9300 BPI (bits per inch) data density. Tape is the oldest memory format still in use, paper punch tapes having gone the way of the stegosaurus . . . to the Smithsonian. Tape has developed into a fairly long-term

and reliable magnetic storage medium. Fresh tapes routinely last as long as ten years with proper care and storage, without loss of data due to print-through, magnetic fading, or spontaneous magnetic reversion. Maximizing the life span of linear computer tape requires not only cool temperatures, sixty-two to sixty-eight degrees Fahrenheit with a tolerable diurnal variation of plus or minus five per cent and a relative humidity (RH) of thirty-five to forty-five per cent and a diurnal variation of plus or minus five per cent, but also occasional cleaning and rewinding at normal tape speed to relieve and redistribute winding tension. Tapes should be stored in the tail-out orientation, and reels maintained vertically in special non-conducting tape racks.² The maintenance of such an environment and care schedule is neither simple nor inexpensive.

Good ANSI (American National Standards Institute) standards exist for computer tape and its storage.³

² Thomas B. Oglesby and William H. Leary, "Managing Electronic Records," NARA Instructional Guide Series (Washington, D.C.: National Archives and Records Administration, 1990), 31. A more detailed technical discussion of the care and maintenance of magnetic recording media can be found in: Sidney B. Geller, "Care and Handling of Computer Magnetic Storage Media," NBS Special Publication 500-101 (Washington, D.C.: National Bureau of Standards and Institute for Computer Science and Technology, 1983).

³ The list of specific standards for the manufacture and testing of magnetic formats is far too extensive to discuss here. The most useful standards are the ANSI/EIA X3 series which deal exclusively with data processing issues.

Manufacturing standards are widely followed throughout the industry. The existence and application of standards by manufacturers mean that computer tape longevity predictions are reliable when the medium is properly stored and protected. The National Archives and Records Administration (NARA) recommends a program of tape replenishment based on a ten-year cycle.⁴

Cartridge tape is a newer and more convenient version of the magnetic reel tape. The format was developed by the 3M corporation during the last decade. The concept is similar to the idea behind the cassette audio tape—the tape cassette is inserted into a streaming tape drive and either a PC hard disk, a file server, or an entire network may be backed up with one or several of these cartridges depending upon the size of the file stored. Streaming tape drives and tapes come in units of ten to eighty megabytes. The tape stock is thinner and narrower than magnetic reel tape and, thus, is more fragile and subject to physical damage, as well as maladies associated with improper storage. Cartridge magnetic tape is more prone to magnetic print-through and linear distortion due to environmental changes and infrequent rewinding or retensioning.

Many users refresh cartridge tapes much more frequently than is necessary for half-inch tapes. General industry agreement is that data should be stable on this format for at least two years.

The optical digital data disk (OD3) is a mass data storage format that has received considerable attention

⁴Oglesby and Leary, "Managing Electronic Records," 30-31.

nationally during the last decade. OD3 systems convert electronic impulses produced and processed by the computer into central processing unit laser pulses which encode a laser-sensitive rotating disk by a variety of methods.

Developing a generally applicable care statement for the optical disk presents some difficulties. There is no single or generic optical disk. The manufacture of these devices and media is a highly competitive, research and development-driven, and proprietary enterprise. Some optical disks are made of tellurium, an extremely unstable rare earth metal, sandwiched between glass plates, and some are constituted of plastic containing embedded polymer compounds that undergo a chemical-kinetic reaction when exposed to laser energy to form a dye spot. As a consequence of this diversity, there are no material or manufacturing standards for optical disks and, because of the newness of the technology, there are no reliable figures on the storage conditions that need to be maintained nor any reliable prediction of the longevity of this recording medium.⁵ Vendors claim a longevity of thirty to one hundred years; all claims are equally groundless and lack scientific verification. With the exception of the glass disks, there is some indication that the OD3 medium is physically durable, it will likely stand up to wear and tear, and it is very resistant to environmental adversity.

⁵ Linda Helgerson, *Introduction to Optical Technology* (Silver Spring, Md.: Association for Information and Image Management, 1987), 15-18.

Optical disks have another and more serious problem which can only be alluded to here. Unlike computer systems, the operating systems of OD3 memory systems are highly proprietary and are held as trade secrets. There is no compatibility standard that will allow one vendor's OD3 system to read another vendor's data disk. This presents a considerable problem for the archivist who must provide access to the optical media in their holdings. With current technological diversity, the archivist is placed in the impractical position of having to accession the system and not just the information medium.⁶

Floppy diskettes, the most familiar magnetic storage medium, have a number of problems from the archival point of view. They suffer from great diversity in the quality of medium available, and there is no prevalent manufacturing or material standard for floppy diskettes that is respected by all manufacturers of the medium. The medium is very prone to environmental damage and is not a very durable or permanent storage format. Neither NARA nor NAC will accept data stored on this transitory medium, nor is the author aware of any state records and archives program that accessions data files on floppy diskettes.

⁶ The issue of data and application portability is under intensive study by both NARA and the National Archives of Canada (NAC). The clearest and most useful document relating to this issue to date is a report of 6 February 1989 (PSC-ARC003-1) prepared by Protocols Standards and Communications, Inc. for NAC entitled, "Application Portability."

Much of the data recorded on floppy diskettes consists of text files which are clearly more appropriately stored for the long-term in human-readable form. While some important databases are stored on floppy disks, it is difficult and may be beyond the technical capacities of the resources of many institutions to convert database files stored on floppy disks into flat or software-independent files for long-term storage and access.

Hard magnetic disks or disk packs are integral parts of a computing systems' internal and rapid access memory (RAM) and will not be considered as an archival medium for electronic data. Only data processing shops regard such devices as storage units for long-term data. And the data processing definition of 'archival' is significantly different from the definition used by archivists. EDP specialists generally use the verb "archive" to indicate the removal of data from the active on-line environment to a storage or inactive environment.

In concluding the discussion of the physical forms and shapes of electronic records, it is necessary to mention the three principal types of electronic records by their functions and structures. This is important because the type of file and its access characteristics will have varied strategies for long-term preservation and reference.

The operational/program file contains sets of machine instructions which tell the CPU (central processing unit) how to deal with and manipulate input and output. An

example of this type of file is DOS for PCs or the application programs sold as dBase or WordPerfect.*

The database file is a collection of data elements assembled to document a similar phenomenon, population, etc., in relatively uniform file structures and composed of like data elements. Database files may be manipulated and reformatted to make certain correlations and relationships more evident. Thus, retaining database files in a machine-readable form, with the proper program, may allow interactive or intelligent inquiry that is highly desirable for the researcher.

There are two ways in which data files can be prepared for archival storage. Data files may be left in software-dependent form and placed in archives along with the version of the software used to structure and manipulate the data files and all supporting documentation. The other alternative, which is somewhat more practical, yet requires more effort to access, is to store the files in a software-independent format, also known as blocked or flat files. These files can either be reconverted to software dependence with programming expertise or may be used in flat form with several statistical analysis packages in wide use. The best known of these packages is SPSS (Statistical Package for the Social Sciences) available for both mainframes and PCs. The PC-based Statistical Analysis Package (SAS/PC) is growing in popularity. These programs work with flat files in either ASCII or EBCDIC

* dBase is a registered trade name owned by Ashton Tate Corporation; WordPerfect is a registered trade name owned by WordPerfect Corporation.

binary formats to restore a portion of an electronic file's interactive capabilities.

The text/document file contains documents created for reading and communications between humans. These documents are created in human-readable language and are stored by the computer in binary code on a magnetic storage device. This file structure is produced primarily by word processing software and electronic mail programs. It is also software dependent and files may be difficult to convert back into a hardware dependent format for use. These files are usually not used for intelligent inquiries; they are simply read, edited, and printed out, or transmitted for reading by others.

Using this information about the primary forms of electronic information storage and the basic types of file structures, it is possible to make some observations about appraising historical records that happen to be in machine-readable form.⁷ It is not possible to consider here the content or historical appraisal process. There are as many standards for appraising records for historical value as there are archivists. It is possible to look at some of the technical factors which affect the decision to accession

⁷ The fundamental work on the influences of the new electronic recordkeeping technologies on historical records is the report prepared for NARA by the National Association of Public Administrators entitled, "The Effects of Electronic Recordkeeping on the Historical Records of the U.S. Government," released in January 1989.

historically significant records that are in machine-readable form.⁸

The physical and access peculiarities of electronic records do make them different from most other types of information formats in institutional archives and influence appraisal decisions. As previously mentioned, there are five general appraisal factors or considerations that may be useful in making appraisal and accessioning decisions about electronic records for an institutional repository. These technical considerations or inquiries will provide the appraisal archivist with direction in determining whether to discard, accession in machine-readable form, or reformat the records series into a more stable, human-readable form. Perhaps it is preferable to think of these appraisal considerations as inquiries which may have a practical affect upon a historical appraisal and accession judgment.

The first inquiry to make about an electronic records series is whether this particular form of the record constitutes the most complete version of the records series that is practical to obtain. This will require data processing knowledge and institutional experience in order to determine if this series exists as a part of another electronic system or exists in another output form, such as hard-copy printout or computer output microform (COM). If the electronic format is the only form of the record series

⁸ An excellent but brief discussion of the National Archives of Canada approach to appraising machine readable records is found in "The Appraisal of Machine Readable Data Files," *Machine Readable Archives Bulletin*, Vol. 1, No. 2 (Summer 1983): 1.

or the form which contains the most complete version of critical historical information, then its preservation must be seriously contemplated. If the most important part(s) of the record series is duplicated or maintained in a more stable format, then the preservation decision will largely rest upon the anticipated type of use the record will receive and the cost of preserving the records series in electronic form.

The second inquiry the appraiser should make about the series is whether it constitutes the most stable form of the record available. Closely related to the question of whether the electronic form constitutes the most complete version of the records series is whether the series or similar record series exist elsewhere in a more stable form such as hard-copy printouts or COM. And, if more stable formats exist, can these records be acquired or are they more likely to be preserved (and available for use)?

The third inquiry relates to the question of whether the electronic record series constitutes the most accessible form of the series, once it is brought into the archives. In other words, does this electronic record series represent the most useful form of the record from the researcher's point of view? A knowledge of the file's initial readability at the time of appraisal and its probable future readability given proper care is fundamental. In addition to investigating its physical integrity, it is essential that the appraiser ascertain whether the series will require the support of such specialized or proprietary software and/or hardware that at the time of accessioning or afterwards it will not be readily accessible to users. Also, if the documentation of a record series or support system taken into the archives is

not complete and clear, then the record series is in danger of losing all accessibility to potential researchers.

Certainly, it is understood that, if the electronic record series is not the most accessible of its alternative forms, then it will not likely be considered for accessioning. Alternatively, if the series in its machine-readable form is the most accessible record series that is available, but not very accessible at the time of appraisal, then series conversion to a more fixed, stable, and human-readable form should be considered prior to accessioning. It is, of course, more justifiable to have a record series that is accessible to users in a non-interactive and human-readable form than to provide to a potential user a format with limited or no accessibility in either human- or machine-readable form.

The fourth inquiry to be made about the electronic records series by the appraiser is whether its current format constitutes the most likely use of the information contained in the records series. More simply, is it more likely that the series will be needed in intelligent and interactive form or will as much or more use be made of the record series in a fixed or non-intelligent format? If the record series is likely to be used for statistical analysis through SPSS, SAS, or other statistical software applications, then as much of the records' value may lie in its format as in its specific content, and accessioning machine-readable formats will be most justifiable. If the records in the series are more likely to be accessed on a single-case basis, then an alternative and fixed format might be considered appropriate. To use the files in this

non-interactive way more closely resembles the use made of textual files than data base files.

It is also useful to remember that many if not most archival customers are trained in the liberal arts and not the methodologies and techniques of the social or mathematical sciences. Machine-readable formats may intimidate many traditional historical researchers, and they will not use series that they might have consulted if available in non-interactive, human-readable forms.

It is also critical to know if the information will have to be sanitized or bracketed before it can be accessed by the researcher. If sanitizing⁹ is required, this may not only affect the series's research value, but also the cost to provide reference service to the series in its machine-readable form. Sanitizing a data file will require the expertise of a data processing specialist and machine time. The need to sanitize a series may render the reformatting of a series and storage in a fixed form more economically justifiable than taking in a more flexible and interactive machine-readable form.

The fifth and final inquiry to be made about the electronic records series under consideration by the

⁹ Sanitizing or bracketing is a data processing term for eliminating specific data fields or data elements from a data base file. Data elements may either be removed from the data file or access to it may be simply prevented on the user's copy. This procedure allows researchers to utilize data bases for statistical analysis which would ordinarily be restricted to all researchers because of privacy rights or other legal restrictions.

appraiser is if its current format constitutes the most cost effective and efficient form of the series that is available. This is perhaps the most cumbersome and the most often disregarded appraisal inquiry. In the case of machine-readable records, this may be one of the most important technical inquiries, as the preservation and access costs of such series can easily drain a small department's budget.

To ascertain a weight for this consideration, the archivist has to devote time to performing some cost calculations and making reference and use predictions for the electronic format and the alternative forms of the series. It is critical to know the annual cost of maintaining the series in electronic form. The cost simply to maintain proper environmental and storage conditions, rewind and recopy data on a scheduled and routine basis is not inconsequential; some authorities estimate the cost for a reel of half-inch computer tape at between eighteen and twenty-six cents per day.¹⁰ This cost must be weighed against the cost to reformat the records series and transfer it to more stable, human-readable, and, in the case of hard-copy printouts, more voluminous, form. COM may be more expensive to produce than paper printouts but will take up significantly less space in the stacks of the archives and probably last considerably longer. The cost to produce

¹⁰ These figures are derived from a paper presented by Walter Meyer zu Erpen of the British Columbia Archives and Records Service at a joint meeting of the Association of British Columbia Archivists and Northwest Archivists in Vancouver, BC, 27 April 1990.

and maintain alternative and more stable forms must be weighed against ongoing maintenance and recopying costs inherent in all machine-readable formats.

In addition to the cost considerations to store and maintain a series in an alternative form, it is important to estimate accurately whether anyone will use or be able to afford to use the series. The computer time and expertise necessary to use, and perhaps sanitize, the records in electronic form is a considerable cost to a potential user. While the cost to store machine-readable formats is not low, the cost to use intelligent records in an interactive manner is considerably more expensive. The cost to copy, sanitize, and process data files when the records series is composed of multiple tapes will relegate such research to researchers holding sizable research grants. The cost of processing such data is falling, but it is not now and will probably not be inconsequential in the near future. In some cases printouts and COM versions of electronic records series may be more accessible and widely available to users than the same record series in electronic form. However, the user's level of creativity and degree of sophistication of analysis may be significantly reduced by the loss of data filed in an interactive and intelligent form.

This article discusses only a few of the questions that are important in the archival and historical appraisal of records in electronic form. It was not intended to be exhaustive, conclusive, or all-inclusive, but merely to consider some issues of fundamental importance in appraising machine-readable records.

Only by examining the practical and ethical questions of reference access and program economics can a balanced

and practical program of electronic record preservation be achieved. Archivists have too often been intimidated by an alien and rapidly-changing technology and by the threat that in the future an increasing quantity of their holdings will be found on electronic formats.

The implication that archivists will be buried under an avalanche of electronic records and that they must be able to care for and provide access to mountainous quantities of this delicate format implies that a critical assumption has been made, a consideration that cannot be taken for granted. The assumption that archivists must accession or anticipate accessioning massive quantities of such exotica as machine-readable records into institutional archives cannot be accepted as axiomatic.

An essential step in determining the proper archival role of machine-readable formats is to determine if archivists can in reality provide them with the storage environments and care which will allow a maximum and a productive life and simultaneously provide potential users with adequate access. This decision is critical and must be evaluated with a full knowledge of the budgetary, technical, and ethical consequences it entails. Archivists and curators must first address the factors inherent in appraising electronic records before accessioning and dedicating themselves and their resources to caring for and feeding these creatures once they are added to the growing archival menagerie.

Michael E. Holland is university archivist at Oregon State University. This article is adapted from a paper given at a meeting of the Society of Southwest Archivists in Austin, Texas, 17 May 1990. The author would

like to thank the Budgets and Institutional Planning Office of Oregon State University for supporting the work involved in delivering this paper to the Society of Southwest Archivists and its subsequent adaption for publication.