9-1-1991

Preservation of Computer-Based and Computer-Generated Records

Gordon B. Neavill
School of Library and Information Studies, University of Alabama, Tuscaloosa, neavill@wayne.edu

Recommended Citation

Neavill, G. B. (1991). Preservation of computer-based and computer-generated records. In K. L. Henderson & W. T. Henderson (Eds.), *Conserving and Preserving Materials in Nonbook Formats* (45-60). Urbana-Champaign: Graduate School of LIbrary and Information Science, University of Illinois.
Available at: http://digitalcommons.wayne.edu/slisfrp/55

GORDON B. NEAVILL

Associate Professor
School of Library and Information Studies
University of Alabama
Tuscaloosa, Alabama

Preservation of Computer-Based and Computer-Generated Records

INTRODUCTION

Half a century after the invention of printing, the German humanist and Benedictine abbot Johannes Trithemius published a book called *In Praise of Scribes* in which he fervently advocated the continued copying of books by hand. He argued that the discipline of copying was spiritually good for the monks, and he pointed out that despite the availability of printed books in large numbers, there remained many unprinted books that were worth copying. But the core of his argument was based squarely on the issue of preservation:

> All of you know the difference between a manuscript and a printed book. The word written on parchment will last a thousand years. The printed word is on paper. How long will it last? The most you can expect a book on paper to survive is two hundred years. Yet, there are many who think they can entrust their words to paper. Only time will tell. (Trithemius, 1974, p. 63)

It is now nearly 500 years since Trithemius wrote, and the era of electronic digital communication is in its infancy. The advent of the computer is comparable in its revolutionary implications to the advent of the printing press. Like printing from movable type, electronic digital communication offers significant advantages over previous technologies, and it has taken hold and spread very rapidly. As the twentieth century draws to a close, more and more records are being created, stored, and disseminated in digital form. A growing number of analog records are being converted to digital form so they can be incorporated into computer-based systems. But just as no one in the fifteenth century fully grasped the implications of the printing revolution, the implications of the computer revolution—social, cultural, economic, and intellectual—remain largely unclear to us.

Especially unclear are the implications for long-term access to digitally encoded information. Issues concerning the preservation and survival of records in digital form have attracted little attention. How long will various forms of digital media last? What steps must be taken to ensure the survival of the information they contain? We have little experience or substantive knowledge to guide policy decisions concerning the preservation of computer-based and computer-generated records, but the implications of what is known are troubling. Many digital storage media have short life expectancies; moreover, retrieval of their information content is dependent on specific software and hardware that may have even shorter life expectancies. Some of the best informed computer experts believe that the only way to ensure long-term access to information is to retain it in human-readable rather than digital form.

Traditional approaches to preservation, with their emphasis on preservation of the storage medium, are largely irrelevant in an electronic environment. The challenge of preserving digitally encoded information for the use of future generations requires rethinking every aspect of the preservation problem in light of digital technologies. This article will raise questions, identify possible problems, and suggest some of the issues that will need to be addressed in the electronic age. Few solutions or specific recommendations are offered. All too often it will be necessary to echo Trithemius's "Only time will tell."

CONCEPTUAL FOUNDATIONS

It will be helpful to begin by trying to establish some conceptual foundations for the problem. Some of the following general statements and conceptual categories apply to the preservation of nondigital media as well as to computer-based and computer-generated records.

The survival of information for archival purposes in any environment depends first of all on the information technology used to capture or transmit the information. Inherent characteristics of the technology largely determine the nature and severity of the preservation problems that will arise. Second in importance are the librarians, archivists, and museum curators who select the recorded messages that will be acquired and retained by their institutions. Third are preservation specialists, most of whose efforts are directed at materials already housed in institutional collections.

Several concepts that illuminate the problem of preservation can be borrowed from communication and information theory. Communication theorists from Harold Innis (1951) to James W. Carey (1989) have explored concepts of space and time. Preservation is concerned with the transmission of information through time rather than space. Some information technologies, such as printing, are effective in transmitting messages through time and space; others are best adapted to one kind of transmission. Many of the information technologies that emerged in the nineteenth and twentieth centuries, such as the telegraph, telephone, radio, and television, are concerned primarily or solely with transmission through space. To be transmitted through time, information normally has to be captured in a documentary record of some kind. Some information technologies do not produce documentary records at all; some produce documentary records that are preservable only with great difficulty; others produce stable documentary records with long life expectancies.

Information theory as it was developed by Claude E. Shannon concerned the transmission of messages through space, but it can easily be applied to transmission through time. Shannon's concept of noise is especially relevant to the preservation of computer-based and computer-generated records. As formulated by Shannon and Weaver (1949), noise consists of additions "to the signal which were not intended by the information source. These unwanted additions may be distortions of sound (in telephony, for example) or static (in radio), or distortions in shape or shading of picture (television), or errors in transmission (telegraphy or facsimile), etc." (p. 99). The degradation of digitally encoded records over time involves the loss of information together with increases in noise that result in reading errors. Error detection and correction codes based on the principle of redundancy mitigate these errors over the short term, but eventually the signal becomes unreadable.

Sidney B. Geller (1974), in a seminal paper that appeared in *Datamation* fifteen years ago, defined information content as *hard* (fixed storage) or *soft* (variable storage). Hard content tends to alter characteristics of the storage medium irreversibly. Print on paper, patterns on film, or coded pits burnt into a substrate by a laser are examples of hard content. Soft content produces changes in the storage medium that are reversible with little or no permanent change in the medium itself. Magnetic technologies in which information content consists of magnetized domains on a ferromagnetic layer are a prime example of soft content.

In an archival system, Geller noted, it is necessary to take into account both *media decay*, the physical deterioration of the medium itself, and *content decay*, the qualitative loss of stored information.

Furthermore, two types of decay rates must be distinguished. *Static decay* is the deterioration of the medium or content as a function of time when it is not being used. *Dynamic decay* is the deterioration of the medium or content as a function of time when it is used in actual operation. Hard content in a medium with a high dynamic decay rate may have a shorter lifetime than soft content in a medium with a lower dynamic decay rate. For example, the hard content of a phonograph record that is played frequently has a shorter lifetime than the soft content of a magnetic tape recording. On the other hand, the phonograph record has a lower static decay rate than the magnetic tape. A rarely played phonograph record should last longer than tape.

Finally, Geller emphasized that, in an electronic environment, thinking about archival quality must "deal with the preservation and security of the entire system which acts as the carrier of the encoded information" (p. 72). Loss of information can result not only from decay of the storage medium but "from the loss of a unique code or decoder (such as a one-of-a-kind transducer)" (p. 72).

DIGITAL TECHNOLOGIES

Inherent Characteristics and Implications for the Survival of Information

The manuscript book that Trithemius trusted would have to be classified in Geller's terms as an example of soft content. Parchment was so tough and durable that the words could be scraped off with a knife. It was also expensive and sometimes hard to come by, so palimpsests—manuscripts written on recycled parchment—were not uncommon. No one knows how much of the lost literature of antiquity and the Middle Ages was destroyed so that parchment could be reused for other purposes.

With magnetic information technologies, we have entered the age of the electronic palimpsest. Magnetic media can be erased and reused with ease. Digital encoding of information coupled with magnetic storage carries this a step further, allowing the erasure of specific data within a record and the interpolation of new data at will. The ease with which data in magnetic digital records can be deleted, modified, updated, and rearranged in new configurations gives computer-based systems a tremendous advantage over print-based systems for the provision of current information. However, it raises serious questions about the survival of information. As this author has written elsewhere, "The malleability of information that is one of the major advantages of computer-based electronic systems has as its corollary the potential transience of information" (Neavill, 1984, p. 77).

The kinds of preservation problems that arise from the malleability of information will be clearer if magnetic digital records are compared with printed records. The information content of printed records is frozen in a particular configuration. In order to revise or correct a printed record, especially if the revisions are at all extensive, a new physical document must be generated. This is normally done by issuing a new edition consisting of multiple copies of the revised text. Copies of the old edition do not cease to exist when the new edition appears. As long as printed documents survive as physical objects, their information content is likely to survive as well.

Ink on paper (especially nonacidic paper) is an effective means of transmitting information through time. A large portion of the paper documents in libraries and archives are outdated or no longer relevant to the purposes for which they were created. Many of these documents are rarely used and remain of interest primarily to historical scholars. The fact that such documents survive as physical objects after they have served the purposes for which they were created is an inherent characteristic of paper-based information technologies. The importance of this characteristic cannot be overestimated. It is largely what makes historical scholarship possible.

In a paper-based environment, outdated city directories, maps, commercial catalogs, membership lists, and publications of many other kinds survive as physical objects and provide important retrospective information to scholars. The value of these records today derives from the fact that their content is frozen. They provide a snapshot of the way things were at a particular time in the past. Successive editions of these publications document and date historical change. For example, a collection of Chicago city directories from 1870 to 1893 would document the economic demography of the city before the Chicago fire, the devastation caused by the fire, and the rebuilding of the city up to the Columbian Exposition.

In a computer-based environment, the content of such information services would be kept continually up-to-date. There is no certainty that outdated information would be retained. Inertia, legal requirements, or its continuing commercial value may cause outdated information to be retained in some systems, as in the Bowker *Books in Print* database, where records of out-of-print books have been retained since 1979. The problem here is that outdated information is likely to be stored in a cumulative file. A cumulative file may contain current and retrospective information, as in the online version of *Books in Print*, or there may be a separate retrospective file, as in the CD-ROM version. A cumulative file provides information about specific out-of-print books, but, unless it is enhanced with costly retrospective features that are irrelevant to the commercial purposes for which the file is created, it destroys the historical context of the information. In contrast, back volumes of the printed version of *Books in Print* can answer the kind of time-specific questions that interest historians.

In an online environment, much of the retrospective documentation that is taken for granted in a print-based environment may be lost. The survival of information is a problem not only where the content of computer-based records is continually revised, but also where computer-based records may be purged from a system because commercial demand has fallen off or because they are no longer relevant to the purposes for which they were created. The loss of historical documentation, or at least the danger of this loss, is inherent in magnetic computer technologies. How serious a problem is this? Is it worth major concern?

The computer is not the first information technology whose introduction has resulted in the loss of retrospective documentation. The invention of the telephone enhanced the ease and speed of spatial communication at the cost of the survival of messages through time. Anyone who has used twentieth-century archives is aware of the dramatic decline in meaningful documentation that resulted from the adoption of the telephone in place of correspondence. Changing technologies of library catalogs provide another example. Until roughly one hundred years ago, most libraries relied on printed book catalogs that appeared occasionally in revised editions. The printed catalog was produced in multiple copies and did a good job of transmitting information about a library's holdings through space as well as time. Its successor, the card catalog, was ineffective on both counts. The compelling advantages of the card catalog were that it allowed the catalog to be kept continually up-to-date, simplified the revision of bibliographic records already in the file, and permitted more access points than were practicable in printed catalogs. In recent years, historians of books have been using printed catalogs to analyze the kinds of books that were available in particular communities in the eighteenth and nineteenth centuries. It will be difficult to carry these studies into the twentieth century. As continually updated files of current information, card and online catalogs cannot provide retrospective documentation about a library's holdings at particular times in the past.

People tend to accept information losses like these, if they think about them at all, because the losses seem unavoidable and because the advantages of the new technologies for millions of users of current information seem to outweigh their disadvantages. This response is rapidly becoming untenable. As more and more of society's records are created, stored, and disseminated in digital form, the level of potential information loss becomes far greater than anything experienced in the past. A recent study that explored problems related to the electronic recordkeeping of government information began its report with the words, "The United States is in danger of losing its memory" (Committee on the Records of Government, 1985, p. 9). It is a question of retaining contact with the past, not simply of providing grist for professional historians' mills. As Edward Shils (1981) has written, "If we could imagine a society in which each generation created all that it used, contemplated, enjoyed, and suffered, we would be imagining a society unlike any which has ever existed" (p. 34). Our inheritance from the past comprises the greatest part of the contemporary stock of knowledge. We draw on that inheritance continually to understand ourselves and our world and to generate new knowledge.

Not all magnetic digital records are subject to continual revision or destruction. Many computer-based systems are designed to collect, store, and manipulate series of historical data such as census and meteorological records. The content of these records may be augmented with new data, but older data are rarely altered or erased. This was not always true in the early years of data processing, when the possibility that future researchers might wish to manipulate data in unforeseen ways was less obvious than it is now. Raw data were not always retained after statistical summaries were prepared. Punched cards containing raw data from the 1940 and 1950 censuses were routinely destroyed, for example (Committee on the Records of Government, 1985, p. 88). Today, the importance of retaining such data is generally recognized. The preservation of digital records whose content is not in flux is not as problematical as the preservation of records whose content is subject to change.

The problems thus far discussed are associated with the soft content of magnetic technologies. They are not inherent in optical technologies, where digitally encoded information is stored as hard content in the form of coded pits burnt into a substrate. Here, information content is frozen in a particular configuration on what appears to be a relatively permanent medium. Optical data discs are written individually; in this respect they can be compared with the manuscript book. CD-ROMs (Compact Disc-Read Only Memory) are produced from a master in editions of multiple identical copies. They are comparable with print media insofar as the survival of information is concerned. When the information content of a compact disc is revised, a new edition is produced; copies of the old edition do not cease to exist.

Currently many libraries are unable to rely on CD-ROMs for retrospective information, but this is a problem that arises from the way they are marketed and is not inherent in CD-ROM technology. Many publications issued in compact disc formats are updated on a regular basis. Each new edition contains the complete database, superseding previous editions. Subscriptions tend to be expensive, so publishers commonly require subscribers to return old discs when a revised edition appears. This precludes the emergence of a secondary market in slightly outdated discs that would cut into the vendor's primary market. It also means that old editions of compact discs, unlike superseded editions of most printed reference works, may not be available in libraries and archives to provide retrospective documentation. There is no assurance that compact disc publishers retain archival copies of outdated discs or that they would permit access to them by scholars if they were retained. In order to fulfill their scholarly responsibilities, research libraries and archives will have to challenge the distribution policies of many compact disc publishers.

Two characteristics of digital technologies enhance the chances for survival of digitally encoded information. First is the ease with which digital records can be copied or converted to new formats without loss of information content. In contrast, information is lost when analog records are copied, and the losses become progressively severe with succeeding generations of copies. Second, although reading errors increase as digital records undergo static or dynamic decay, error detection and correction devices in digital records can mitigate information losses even in records that are in advanced stages of degradation. End-of-life of digital records is reached only when errors overwhelm the ability of error detection and correction devices to correct them. Without these two characteristics, it is doubtful whether digital media could be considered for purposes of transmitting information through time.

Preservation of Digital Storage Media

Two distinct technologies, magnetic and optical, are used for the storage of digitally encoded information. Data tape, hard discs, and floppy discs are examples of magnetic media. Optical data discs and compact discs are examples of optical media. Erasable magneto-optic discs combine magnetic and optical technologies.

The lifetime of magnetic storage media is measured at best in decades; optical storage media may have a somewhat longer life expectancy. In contrast, the lifetime of archival-quality paper and photographic film is measured in centuries. Digital storage media are not considered to be of archival quality (National Research Council, 1986, pp. 68, 76). In the long run, the preservation of digitally encoded records is concerned with the preservation of content rather than storage media. The preservation of digital storage media focuses on the short term, generally ten to twenty years. Here, the objective is to maintain information content until it can be copied or until the file is converted to a new format. Preservation at this level is concerned with identifying the characteristics of digital storage media that contribute to degradation, establishing conditions for storage and handling, monitoring the deterioration of storage media, and scheduling copying or file conversion. Most of what is known about the preservation of digital storage media relates to magnetic data tape. The following discussion focuses on this medium, with some tentative comments about optical media.

Magnetic data tape is subject to both dynamic and static decay. Since the tape is in direct contact with the heads of the transport when it is written and read, a certain level of dynamic decay is unavoidable. The primary problem, however, is static decay. To understand the nature of the problem, one needs to look at the composition of the tape.

Magnetic tape consists of a ferromagnetic layer bonded to a polyester base. The base is stable against environmental degradation under ordinary use (Cuddihy, 1980, p. 558). A National Bureau of Standards study found that the base, stored at 20°C and 50 percent relative humidity, should retain useful properties for 500 years (Smith, 1986, p. 2). The ferromagnetic layer is composed of ferromagnetic particles of about one micron in length. The useful properties of the tape derive from these particles. Information on the tape is stored and manipulated in terms of the magnetic qualities of these particles, whether they are magnetized or not, and the direction of the polarities. The particles themselves are chemically stable. The problem lies in the binder in which the particles are suspended. Binder formulas vary from one tape manufacturer to another, but they consist basically of polyester polyurethane to which various lubricants, adhesives, and stabilizers (against such dangers as oxidation or bacterial growth) are added. The binder is subject to environmental degradation by hydrolysis, a chemical reaction between the polyester polyurethane and atmospheric water vapor. High temperature, high humidity, and acidic pollutants in the air accelerate the hydrolytic reaction. In a new tape, the binder has a high force of adhesion to the base. As the tape ages, the adhesion drops rapidly and the binder layer becomes brittle. The weakened binder may flake off the base, and a gummy substance may be exuded. The gummy substance can lead to increased error rates when the tape is read (Bertram & Cuddihy, 1982, p. 993). Tape lifetime is limited by degradation and detachment of the binder from the base.

The National Bureau of Standards recently concluded a five- or six-year laboratory study titled "Prediction of the Long Term Stability of Polyester-based Recording Media." Data were written on tapes which were rapidly aged at known temperatures and relative humidities. Data were read after aging. Aging and reading attempts then alternated until the tapes became unreadable. The study concludes that tape lifetime is easily ten to twenty years, probably longer, under normal storage conditions and room temperature and 50 percent relative humidity. However, the authors cautioned, "There are documented reports of tape failure after ten years of storage under normal room temperature and humidity and we have seen cases of failure after only a few years so the lifetimes can vary considerably" (Smith et al., 1986, p. 1). No systematic differences were found between the commercially available tapes tested. Tapes sold as "archival" at higher prices apparently last no longer than other tapes, although they may have superior performance in other respects (L. E. Smith, personal communication, October 25, 1988).

Modern tape transports can retrieve information from digital tapes in terrible condition, including very brittle tapes, but there is an endpoint at which tapes can no longer be read. The only way to preserve the information content of magnetic data tapes is to recopy them on a regular basis, perhaps every twelve to twenty years. One of the objectives of the National Bureau of Standards study was to develop a simple, qualitative test for binder separation to help determine when a tape needs to be copied. The simplest test is apparently to crease an insignificant part of the tape and roll it between thumb and forefinger. Aged tapes often lost binder, sometimes when creased and sometimes after several passes between the fingers (Smith, 1988).

A number of guidelines are commonly suggested for tape storage and maintenance. These include sampling tapes on a regular basis for errors, checking for deterioration, cleaning and rewinding tapes every one or two years, and eventually, recopying. It is essential to create an identical copy of the original tape, preferably on tape of a different manufacturer, and to store it under good conditions in a separate facility. The National Bureau of Standards publication, *Care and Handling of Computer Magnetic Storage Media* (Geller, 1983), provides a thorough summary of current practices. There is also a useful brief article by Benjamin L. DeWhitt of the National Archives and Records Administration in *Conservation Administration News* (DeWhitt, 1987).

Floppy discs have a somewhat shorter life expectancy than magnetic tape. The National Bureau of Standards estimates that a floppy disc stored under proper conditions can provide information for ten to fifteen years; archivists estimate that information on a floppy disc may last no more than five years (Committee on the Records of Government, 1985, p. 32). Many people store correspondence files, original versions of published writings, and other personal archives on floppy discs. These archives must be copied or converted to new formats on a regular basis. Otherwise, they are likely to be lost or become inaccessible because of dependence on obsolete hardware or software.

Optical technologies are far more stable than magnetic technologies. Here we are dealing with hard content: coded pits burnt into a substrate of glass, metal, or durable plastic. The information content is read by a laser so there is no physical contact between the reading head and the surface of the disc and virtually no basis for dynamic decay. There are, however, grounds for concern about static decay. The National Research Council's Committee on the Preservation of Historic Records (1986) has noted,

> The factors of storage that are most harmful to the optical disk are heat and humidity. Humidity causes oxidation of the recording surface, and heat accelerates the process. Some of the other factors that cause concern are the adhesion of various layers one to another, catalytic corrosion, galvanic corrosion, and mechanical stresses. (pp. 73-74)

It is sometimes asserted that compact discs are virtually indestructible. One author has written that you could probably use a compact disk "to play Frisbee with your dog" (Miller, 1986, p. 22). This is fantasy; optical media have to be handled carefully. Manufacturers tend to be more conservative. Many vendors claim a life of ten to forty years for CD-ROMs, with most of the claims on the low end of that range. These claims do not appear to be based on test results or other reliable data. It may simply be a case of caution on the part of vendors who have no idea how long their products will last.

The durability of optical media is related to the materials from which the discs are made. Discs using a glass substrate with a gold alloy or platinum coating layer are the most durable. Commercially produced CD-ROMs, which use a plastic (polycarbonate) substrate with a chemically reactive coating layer such as aluminum, are the least durable. The Library of Congress has conducted accelerated aging tests of optical discs. Glass and gold alloy discs have been baked for 2,000 hours without bringing them to end-of-life. Discs are now being tested after having been baked for 3,000 hours (B. Nugent, personal communication, October 1988).

System as a Whole

The fact that digital storage media are not of archival quality is unimportant in terms of preservation. Geller (1974) noted that the term "lifetime" applies to storage media, information content, and the system, which he defined as "the total grouping of all of the interrelated components, the storage medium and the encoding/decoding schemes"

(p. 75). Little is gained if storage media outlive other components of the system, including hardware and software, that are required to provide access to their information content.

The crucial factor in the preservation of computer-based and computer-generated records is the life expectancy of the system as a whole. According to Geller, system end-of-life can be caused by obsolescence due to technological or sociological advances that make maintenance of an existing system costly and inefficient, inability of the system to perform reliably after a certain time, loss of one-of-a-kind transducers or encoding/decoding keys, or a breach in a highly secure system (p. 75). The following discussion will focus on the third category.

Electronic hardware is not expected to function for more than ten to twenty years. Generations of incompatible hardware rapidly succeed each other. When computer tapes containing raw data from the 1960 federal census came to the attention of the National Archives and Records Service in the mid-1970s, only two machines in the world were capable of reading them. One was in Japan; the other was deposited in the Smithsonian as an historic relic (Committee on the Records of Government, 1985, pp. 86-87). Successive generations of discontinued hardware cannot be maintained by manufacturers or archival institutions to provide access to superseded digital formats.

Software is also necessary for the recovery of digitally encoded data. Software tends to change more rapidly than hardware, and generations of software tend to be even less compatible than generations of hardware. There are examples of software that have gone through thirty versions in a decade. The proper operating system must be available at the time of data recovery.

Another problem is documentation of the software. Not only must the proper operating system be available, but documentation is necessary to give information about the digital codes used and the organization or format of the record. Documentation may be in machine-readable or human-readable form. It is not unusual for documentation to be incomplete or missing.

Security is a preservation issue that deserves to be dealt with at greater length than is possible here. In order to protect data, it may be necessary to encode them in more cryptic form. Higher density storage also requires more sophisticated coding and decoding schemes. In both cases, recovery of data becomes more difficult if the key to the code is lost. And, of course, the amount of information that can be lost if anything goes wrong increases with the density of storage.

The long-term preservation of computer-based and computer-generated records requires constant vigilance on the part of the archivist. Records must be copied on a regular basis to prevent loss of information due to media or content decay. In practice, most copying will involve conversion to new formats to ensure that the records remain compatible with current systems. Original software must also be preserved and copied. Documentation must be carefully preserved, ideally with the records themselves. The costs of this kind of preservation are high, but if these procedures are carried out conscientiously, they can ensure "an almost endless data lifetime" (Geller, 1974, p. 76).

The risks are high as well as the costs. In 1984, a Committee on Preservation established by the National Archives and Records Service to advise on machine-readable records issued a white paper that argued strongly against the preservation of archival records in machine-readable form (National Archives and Records Service, 1984; Mallinson, 1985b). The committee members had impressive technical credentials and represented such institutions as the Center for Magnetic Recording Research at the University of California, San Diego; IBM; the 3M Corporation; the National Security Agency; and the National Bureau of Standards. Their opposition to machine-readable archives was based primarily on problems related to hardware and software dependence. They proposed, instead, a system relying on archival quality silver halide microfilm. The advantage of microfilm mass memory is that the information content is permanently stored in human-readable form. Access to the information content merely requires magnification, not complicated decoding devices. Members of the committee met again in the summer of 1988. Despite advances in optical storage technologies, they remain strongly convinced that human-readable storage is the only kind that makes sense in an archival situation (Smith, 1988). In their view, machine-readable storage for preservation purposes is simply too risky, and it condemns the archival agency to perpetual copying and reformatting of records.

POLICY ISSUES

The preservation of computer-based and computer-generated records for the use of future generations is one of the greatest challenges facing librarians, archivists, records managers, and other members of the information community. We are just beginning to understand the dimensions of the problem. The next step will be to formulate information policies to ensure the survival of digitally encoded information. A few of the policy issues that will have to be addressed are outlined below.

The first issue to be resolved is whether records are to be retained in human- or machine-readable form. Although there are strong arguments against converting existing archival records to machine-readable form, it is unlikely that records created and disseminated in digital form will be widely converted to human-readable form for preservation purposes. Records created specifically for manipulation by machine, such as interactive fiction, hypertext, and statistical and bibliographical data files, must be retained in machine-readable form. Digital storage may not be advisable in all cases, but the problem of preserving machine-readable records cannot be avoided.

The next issue concerns the locus of responsibility for preservation of computer-based and computer-generated records. As stated earlier, the survival of information for archival purposes in any environment depends, first of all, on the information technology used to capture or transmit the information and secondly, on the librarians, archivists, and museum curators who select the recorded messages that will be acquired and retained by their institutions. Commercial and other organizations that generate and disseminate information cannot be expected to assume an archival role. Preservation is not their responsibility; in any case, they go out of business or change their policies too frequently to be reliable. If computer-based and computer-generated records are to be preserved, they will have to be preserved in institutions dedicated to that purpose.

The selection of records to be acquired for preservation purposes will be far more difficult in an electronic environment. Decisions to acquire computer-based records will have to be made before their information content is erased. Archival institutions in a print-based environment can acquire records years after their creation, when their research significance is established. Digitally encoded information that is not acquired when it is current may be lost forever. The librarians and archivists who select computer-based records for acquisition and retention will have to anticipate the needs and interests of future generations as best they can. A national information policy to ensure that essential information is preserved would be desirable.

The preservation of continually updated databases for purposes of retrospective documentation is an especially difficult problem. The best solution seems to be to download information from the database at established intervals which would provide periodic snapshots of the database with the information content frozen at a particular moment in time. The snapshots could be in whichever format is eventually accepted for archival storage. The critical issue for future scholars would be the selection of databases to be preserved.

The acquisition of computer-based records by institutions dedicated to preservation raises economic and legal issues that will have to be addressed at the national level. Fewer institutions will need to assume an archival role in an electronic environment, but responsibilities for preservation will have to be coordinated to ensure that essential records are preserved. Patterns of economic support for research libraries will have to be reordered. In a print-based environment, current publications constitute the vast majority of acquisitions at most research libraries; retrospective information needs are served in large part by materials whose original purchase was justified on the basis of their provision of current information. The economic link between the provision of current and retrospective information is broken in an electronic environment where libraries provide access to a wide range of computer-based records they do not own. Economic support specifically for the acquisition and preservation of retrospective materials will have to be increased.

CONCLUSION

As we grapple with these issues in the years ahead, we will have come full circle to Trithemius. He was in no sense opposed to the new technology of printing. He fully recognized the advantages of printing for disseminating ideas. His own books, including *In Praise of Scribes*, were printed, and his circle kept one Mainz printshop so busy that it practically became the abbey press. He was wrong about the life expectancy of the fifteenth-century printed book, but he grasped that a single technology may not be equally effective in transmitting information through both space and time. Our thinking for the past half millennium has been the product of a world where the dominant information technology accomplished both objectives fairly well. As we enter the electronic era, we must try to think again of transmission through space and time as separate issues. Assumptions derived from a print-based environment about permanence and the cumulative growth of knowledge may give way to a heightened awareness of the fragility and vulnerability of the stock of human knowledge, and a return to the unending medieval preoccupation with preservation and recovery.

REFERENCES

Bertram, H. N., & Cuddihy, E. F. (1982). Kinetics of the humid aging of magnetic recording tape. *IEEE Transactions on Magnetics, 18*(5), 993-999.

Carey, J. W. (1989). *Communication and culture: Essays on media and society*. Boston, MA: Unwin Hyman.

Committee on the Records of Government. (1985). *Report of the Committee on the Records of the Government.* Washington, DC: The Committee.

Cuddihy, E. . (1980). Aging of magnetic recording tape. *IEEE Transactions on Magnetics, 16*(4), 558-568.

DeWhitt, B. L. (1987). Long-term preservation of data on computer magnetic media. *Conservation Administration News, 29*(April), 7, 19, 28; *30*(July), 4, 24.

Geller, S. B. (1974). Archival data storage. *Datamation, 20* (October), 72-80.

Geller, S. B. (1983). *Care and handling of computer magnetic storage media.* NBS Special Publication 500-101. Washington, DC: National Bureau of Standards.

Innis, H. A. (1951). *The bias of communication.* Toronto: University of Toronto Press.

Mallinson, J.C. (1985a). The next decade in magnetic recording. *IEEE Transactions on Magnetics, 21*(3), 1217-1220.

Mallinson, J. C. (1985b). Archiving human and machine readable records for the millennia. In *Second international symposium: The stability and preservation of photographic images* (August 25-28, 1985) (pp. 60-63). Springfield, VA: Society of Photographic Scientists and Engineers.

Miller, D. C. (1986). Finally it works: Now it must "Play in Peoria." In S. Lambert & S. Ropiequet, (Eds.), *CD-ROM: The new papyrus* (pp.21-35). Redmond, WA: Microsoft Press.

National Archives and Records Service, Committee on Preservation, Subcommittee C. (1984). *Strategic technology considerations relative to the preservation and storage of human and machine readable records.* White Paper. Available from NARS, Washington, DC.

National Research Council. Committee on Preservation of Historical Records. (1986). *Preservation of historical records.* Washington, DC: National Academy Press.

Neavill, G.B. (1984). Electronic publishing, libraries, and the survival of information. *Library Resources & Technical Services, 28*(1), 76-89.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Shils, E. (1981). *Tradition.* Chicago: University of Chicago Press.

Smith, L. E., (1986). *Prediction of the long-term stability of polyester-based recording media: Progress report to June 1986.* NBSIR 86-3474. Gaithersburg, MD: National Bureau of Standards, Institute for Materials Science and Engineering, Polymers Division.

Trithemius, J. (1974). *In praise of scribes.* (K.Arnold, Ed.; R. Behrendt, Trans.). Lawrence, KS: Coronado Press.