Counterpoint

Preserving Machine-Readable Archival Records for the Millenia

by **JOHN C. MALLINSON***

Machine-readable records such as magnetic tapes and optical discs, unlike photographs or maps, which can be examined by the naked eye, can only be read by putting them in a machine. This fact introduces a number of novel problems for archivists. Some of these problems have been discussed in a White Paper which Subcommittee C of the Committee on Preservation of the National Archives and Records Administration presented to the Archivist of the United States in July 1984. The title of the White Paper is "Strategic Technology Consideration Relative to the Preservation and Storage of Human and Machine Readable Records." The subcommittee was asked to advise the National Archives and Records Administration on the long-term prospects for preservation of machine-readable records. The White Paper concludes that human-readable microfilm provides the simplest and most effective means of permanent retention of the information in these records. (Microfilm was chosen because there are only two archivally certifiable recording media extant — paper and microfilm.) Looking back, it took an enormous length of time, five years, to come to what seems now to be a very obvious and simple conclusion.[1]

* This paper was originally presented at the Second International Symposium: The Stability and Preservation of Photographic Images, at the Public Archives of Canada in Ottawa in August 1985. The symposium was sponsored by the Society of Photographic Scientists and Engineers.

1 All the members of the subcommittee, myself, John Davis, Walter O'Neill, Andrew Persoon, Leslie Smith, and Richard Zech, are technologists. None are archivists or historians. They are all associated with what is now called the "Information Revolution," that is to say, electronic data transfer, electronic recording of information, electronic transmission, and all the things that one associates with computers, satellites, and electronics.

Subcommittee C of the Committee on Preservation was formed in 1979. I was the chairman; at that time I was the Manager of Research at the Ampex Corporation, which invented video recording. Ampex also was the first company in the Western World to commercialize magnetic recording. I am a physicist. John Davis is the Chief of Storage Research at the National Security Agency. (His basement has been

In order to come to grips with this preservation problem, Subcommittee C had first of all to decide what were the actual priorities operating in an organization such as a national archives. It concluded that the first priority, the top priority, was to accession and preserve valuable records indefinitely and make them available to the public. The key word is "indefinitely." It was the subcommittee's understanding that the mission of a national archives is to accession material and retain it indefinitely. It will never be thrown away. It is the eternal record, and, in making it available to the public, the subcommittee understood that there is no great urgency. If the record can be made available in a few days, that is fine. The principal mission is to retain the information indefinitely. The second priority is to make those records in the archives available in an easy, timely, and inexpensive manner. By timely, we imagine that if records can be made available in five or ten minutes, that is acceptable. The third priority, the goodwill priority of a national archives, is to supply archival records, or the information in them, in an easy-to-use or electronic form or provide data processing services. This is the activity that perhaps causes the most trouble in the archivist's mind, because this priority is the one that necessitates the computer-driven environment. This requires a central computer driven by magnetic tapes, or perhaps an optical disc. Whatever the principal store is, it is transmitting images; it is quickly disseminating information to a multiplicity of users. It is a broadcast system. The users may perhaps be at a hundred video terminals scattered around the building. The data may be transmitted via satellite across the breadth of the nation.

When the subcommittee first started to consider the preservation problem of machine-readable records, it perceived incorrectly that it had to advise the National Archives how best to accession half-inch computer tape, because that is the predominant machine-readable medium today. Half-inch computer tape is today's medium of exchange between computing centres. The subcommittee spent a long time wondering whether a videotape or digital magnetic image will be readable one hundred years from now when stored at reasonable ambient conditions. The subcommittee slowly but surely came to the conclusion that the answer to the question was moot, or almost irrelevant. One hundred years from now, no one will know the answer because other factors are operative, the most critical of which is the short lifetime of the machine itself.

It is a reasonable assumption that most current magnetic tape and most of the optical discs that are being made today can be expected, if given reasonable care, to last for twenty or perhaps thirty years. Whether it is twenty or thirty years, or whether it is fifty years, is almost irrelevant because the subcommittee considers it to be a certainty that the machines that will be required to read those tapes in thirty or forty or fifty years will not be available and serviceable. The problem with machine-readable records is the long-term availability of the machines rather than the physical decay of the recording medium. Of course, that is a revolutionary thought in the archival world which, traditionally, has

---

filled with every kind of optical film and disc recorder known to mankind over the last fifteen years.) Walter O'Neill is the Special Assistant for Federal Policy in IBM's Government Systems Division. Andy Persoon is now retired, but he was the Technical Director of the Minnesota Mining & Manufacturing Company and is an expert on audio and video tape. Leslie Smith is at the U.S. National Bureau of Standards. He is an organic chemist in the polymer science and standards division. He has written many papers, some under contract for the National Archives in Washington, on the subject of hydrolysis of the polyester base film of tape and film and the polyester-urethane binder systems in tape. Richard Zech, who was in fact the author of the White Paper, was at that time Vice-President of the Electronic Publishing Department of McGraw-Hill, and is now a Vice-President of his own company, ISI, in Colorado Springs. ISI will be manufacturing optical discs and optical disc recorders.

been concerned with such questions as how fast is this paper deteriorating; why is the vellum doing this, that or the other; why is the colour fading in the colour photographs; what is the archival life of a photograph? Those questions arise in regard to all human-readable records. Machine-readable records raise another class of question, the most critical one of which is how long is the machine going to last.

There are three basic problems with machine-readable records. As a technologist, it hurts to have to confess that electronics has these weaknesses. VAX computers, IBM PC computers, satellite relays, and broadcast networks are in place because they satisfy the human need for very rapid transfer of information. But that rapid transfer of information has a tremendous price. In order to achieve the greater data speed and storage capacity, all storage densities are increasing. This trend has been evident across this century. Compare the storage density of an Edison wax cylinder, a 78 RPM phonograph record, a 33 RPM record, and an optical compact disc. The storage densities are increasing because they deliver greater speed and greater capacity.

Any trend towards higher storage density is fundamentally flawed if long-term archival storage is intended. It does not take a physical chemist to grasp that very, very small features are more prone to decay or degradation than large features. There are conflicting interests. The electronics world wants higher densities; the archival world requires lower densities.

The compact audio disc optical recorder is a commercial success. Patty Two is the highest density magnetic disc file made in Japan by the Japanese equivalent of the Western Electric Company, NTT, the Nippon Telegraph Company. The IBM 3385 is the top of the line IBM mainframe disc file, a $100,000 machine. The new format for half-inch tape is eighteen bits, two bytes wide, 19,000 BPI in a little cassette; it is used in the IBM 3840. It has the same capacity as the old half-inch tapes, the 1600 or 6250 BPI tapes, but is much more compact. Eight millimeter video is the emerging new home VCR standard.

The bits per square inch range from the old half-inch tape, which is .03 million bits to the square inch (very, very low) all the way up to, say, the compact audio disc or optical read-only-memory disc at 400 million bits to the square inch.

It is frequently the case in reports written by archivists that a comparison is made between this old half-inch tape, which was a standard that was started in about 1970, and the new audio compact disc. When you compare 400 with .03 you get a ratio of 12,000 to 1. Thus, it is said that optical disc technology affords an extremely large increase in storage density.

If, on the other hand, the highest density magnetic recording product at the moment, the Patty Two, is taken as the base line, then the compact audio disc is only a factor of ten higher in area density. If the emerging eight millimeter home video recorder is the base line unity, then there is only a factor of five difference. Area densities are increasing inexorably and, give or take a factor of five or perhaps ten, there is not that much difference between optical and magnetic recording.

The second problem is that the pace of technological change is such that the expected or planned useful life of electronic devices is less than twenty years. I know of no electronic device which has lasted and been in common commercial use for more than twenty years. Ten years is quite often a reasonable target. Even when scientists build communication relay satellites that have been assembled in clean rooms at costs of literally millions of dollars, a ten-year life span seems to be unattainable. Within twenty years a manufacturer is almost certain to abandon maintenance of a machine. For example, the Digital Equipment Corporation, which makes PDP-11 and VAX computers, is not servicing the PDP computers anymore. At Ampex, a PDP-11 was declared to be irreparable by the manufacturer when it was eleven years old. Today's popular IBM PC Junior will have the same fate.

Once upon a time, for instance, a 78 RPM phonograph player could be used even if the original needles had run out. As a boy, I used pins. It does not require much of a machine shop to mend a clockwork motor or a simple little electric motor to keep the record turning. There were no electronics involved. The player could be maintained indefinitely. That is not so today. The video heads in all video recorders are limited life items. They are certainly not reparable now. Integrated circuits are never repaired. They are considered by the electronics industry to be irreparable. Laser diodes used in optical disc recorders cannot even be exchanged anywhere in the United States. They all go back to Japan to be exchanged.

A home video recorder, a VHS recorder, whatever the purchase price or brand, is going to wear out after about two thousand operating hours. Two thousand operating hours may seem a very short life, but a car is probably going to wear out after two thousand operating hours. Most machines have a lifetime of a few thousand hours.

Professional video recorders have a mean time before failure (MTBF) of about two thousand hours. For half-inch computer tape drives, IBM says mean time before failure is five thousand hours. Fourteen-inch disc files have a MTBF of eight thousand hours; a remarkable technological feat. With audio compact disc players, the expected life of the gallium arsenide laser diode is between one thousand and fifteen hundred hours. The mean time to repair is very short. All of these machines are arranged so that new heads or new optical heads can be put in and aligned quite quickly.

However, the manufacturer's support may be withdrawn for a particular piece of hardware. If Sony says optical heads are no longer available or Ampex says video heads are no longer available, the chances are that this machine is going to malfunction within two or three years. And then it will be dead in the water. It will be useless.

It is in this context that Subcommittee C began to realize that all the discussion about whether an optical disc will last for twenty years or thirty years, or magnetic tape one hundred years becomes irrelevant if the machine cannot be maintained once the manufacturer withdraws its support.

All modern machines have certain critical components which cannot be repaired. A good example is an integrated circuit or a microprocessor chip. When the manufacturer makes them at the moment, they are tested, and defective ones are thrown away. The manufacturer makes no attempt to repair them. They can only be fabricated in special multi-million dollar factories. They cannot be substituted from a different machine. The video head cannot be taken out of a betamax recorder and put it into a VHS recorder. They are not interchangeable.

The underlying reason, of course, for these problems is the tiny size of these components. The track width of a video head is now seven-tenths of a thousandth of an inch. The large scale integrated circuit connection wires or traces are only one or two microns wide.

The day has long since passed when a competent technician in a well-equipped workshop can keep things going.

A further problem is the number of different formats each medium may have during its history. Video recording was invented in 1956, and since then there have been no less than eight different video tape formats. They are: the two-inch quad; the double density two-inch quad; the type A helical, the type B helical, the type C helical, which is the one used in the broadcast industry today; the three-quarter inch u-matic; the half-inch betamax; the half-inch VHS; and the eight-millimeter. This means that on average the format is changing every three or four years. Each one of these tape formats requires a different machine to play it. They are not backwardly compatible. In half-inch computer tape, the medium of exchange between computers since 1952, there have been eight different half-inch tape formats. The list includes: 100 BPI-7 track; 200 BPI-7 track; 560 BPI-7 track; 800 BPI-7 track; 1600 BPI-9 track; 3200 BPI-9 track; 6250 BPI-9 track; 19000 BPI-18 track.

Once again, one comes to the conclusion that every four or five years a new tape format is adopted. The computer industry has been better than the video industry in that some of these are backwardly compatible. Most of the 9-track tapes will play on previous 9-track machines, but clearly a 9-track tape will not play on a 7-track machine or vice versa. In fact, the IBM Corporation a few years ago announced that it was no longer supporting 7-track recorders. There are some records in the Department of Defense in Washington, D.C. which are now unusable, not because the tape has deteriorated, but because the manufacturer will not support the machine anymore. The message is clear that both in video and computer recording, the format changes every four or five years.

In optical disc recording, it is fair to say that the world is now on its third generation of optical discs. There have been analog video discs from Philips, analog video discs from Pioneer, and, now, the digital audio compact discs.

For the above reasons Subcommittee C concluded that by using a human-readable microfilm mass memory approach, the impossible task of trying to maintain obsolete machines can be avoided. This approach guarantees the indefinite preservation of archival records.

Now, it may be argued at this point that, clearly, records can be maintained indefinitely in computers. The Internal Revenue Service or the equivalent in Canada never makes mistakes. It is true that, with digitally recorded information, it is possible to perform error detection and correction and indefinitely regenerate a perfect message. This should come as no surprise to archivists and historians because all human writing is essentially digital in nature. Digital simply means that instead of a continuum of things to be recorded such as the gray scale in a photograph, it is agreed in advance to use only a limited lexicon, the twenty-six letters in the alphabet.

Consequently, if my text contains a spelling mistake, it can be seen and corrected very quickly. Similarly, in binary digital recording, where ones and zeros only are being recorded, it is possible to add some redundancy and make the corrections. So it is possible to keep on regenerating the information indefinitely in digital recording and it would indeed be possible, as is routinely done in nearly every computer centre, to keep on doing file conversions, that is to say, taking the old information, correcting it so that it is perfect, and re-recording it in the new format on the new machine.

But the question Subcommittee C had to think about was, can one really imagine that, as computers and electronic data processing become more and more prevalent in the future, a national archives, or any other archives, could seriously undertake a plan of re-recording the information, say, every ten years. It does not require a mathematician to see the consequences of that. Very rapidly the point will be reached where the whole archive will be re-recorded every ten years. Subcommittee C decided that such massive file conversion was prohibitively expensive. Thus it recommended human-readable microfilm, which guarantees indefinite preservation, in the old-fashioned human-readable sense and not in the modern error detection, correct, and re-recorded sense. It provides a hardware and software independent mode of preservation. It is amenable to mechanization, that is it can be surrounded with all the glories of technology. It is risk-free. It provides a basis for a totally integrated modern records management and preservation system.

Machine-readable records need machines. Without them they are useless. The machines are the principal source of the archival problem, not the records, tapes, or the discs. And it seems unlikely that future machines will solve the problem. Indeed, it seems more likely that they will exaggerate rather than solve the problem.