A LA RECHERCHE DU TEMPS PERDU: ETERNITY IN CYBERSPACE

by M. E. Kabay, PhD, CISSP
Security Leader, INFOSEC Group, AtomicTangerine, Inc.

Everyone makes backups and stores them, right? And everyone keeps archives of electronic data in accordance with legal requirements or organizational policy, right?

Well, no.

Many of us are storing records in ways that make it unlikely we will ever be able to read them in the long term required for archival use. And archivists ought to know better.

Storing records is only half the task of records management; supporting availability and utility is the essential function. No one wants a WOM (write-only memory) for their records. For short-term storage, there is no problem ensuring that stored information will be usable. Even if a software upgrade changes file formats, the previous versions are usually readable. In a year, technological changes such as new storage formats will not make older formats unreadable.

Over the medium term, up to five years, difficulties of compatibility do increase, although not catastrophically. There are certainly plenty of five-year old systems still in use, and it is unlikely that this level of technological inertia will be seriously reduced in the future.

Over the longer term, however, there are serious problems to overcome in maintaining the availability of electronic records. Over the last ten to twenty years, certain forms of storage have become essentially unusable. As an example, AES was a powerful force in the dedicated word-processor market in the 1970s; eight-inch disks held dozens or hundreds of pages of text and could be read in almost any office in North America. By the late 1980s, AES had succumbed to word-processing packages running on general-purpose computers; by 1990, the last Canadian company supporting AES equipment closed its doors in Montreal. Today, it would be extremely difficult to recover data from AES diskettes.

The problems of obsolescence include data degradation, software incompatibilities and hardware incompatibilities.

Magnetic media degrade over time. Over a period of a few years, thermal disruption of magnetic domains gradually blurs the boundaries of the magnetized areas, making it harder for I/O devices to distinguish between the domains representing ones and those representing zeroes. These problems affect tapes, diskettes and magnetic disks and cause increasing parity errors. Specialized equipment and software can compensate for these errors and recover most of the data on such old media.

Tape media suffer from an additional source of degradation: the metal oxide becomes friable and begins to flake off the Mylar backing. Such losses are unrecoverable. They occur within a few years in media stored under inadequate environmental controls and within five to ten years for properly-maintained media. Regular regeneration by copying the data before the underlying medium disintegrates prevents data losses.

Optical disks, which use laser beams to etch bubbles in the substrate, are much more stable than magnetic media. Because CD-ROMs and laser disks are still so new, no one knows how long optical disks will last; nonetheless, technologists predict that the information will remain readable for decades and more. They will remain readable if and only if future CD-ROM systems include backward compatibility.

Software incompatibilities include the application software and the operating system.

The data may be readable, but will they be usable? Manufacturers provide backward compatibility, but there are limits. WordPerfect 9 can convert files from earlier versions of WordPerfect--but only back to version 4.2. Over time, application programs evolve and drop support of the earliest data formats. Database programs, E-mail, spreadsheets--all of today's and tomorrow's versions may have trouble interpreting data files correctly.

In any case, all conversion raises the possibility of data loss since new formats are not necessarily supersets of old formats. For example, in 1972, RUNOFF text files on mainframe systems included instructions to pause a daisy-wheel impact printer so the operator could change daisy wheels--but there was no requirement to document the desired daisy wheel. The operator made the choice. What would document conversion do with that instruction?

Even operating systems evolve. Programs intended for the DOS of a decade ago do not necessarily function on today's DOS version 6.20. And the operating systems of yesteryear do not necessarily run on today's hardware. Even emulators can cause problems because, again, there is no guarantee of compatibility between the emulated system and the emulator.

Finally, even hardware eventually becomes impossible to maintain. As mentioned above, it would be extremely difficult to retrieve and interpret data from word-processing equipment from even twenty years ago. No one outside museums or hobbyists can read an 800 bpi 9-track ¾-inch magnetic tape from a 1980 HP3000 Series III minicomputer. Over time, even such parameters as electrical power attributes may change, making obsolete equipment difficult to run even if they can be located.

The most robust method developed to date for long-term storage of data is COM (Computer Output to Microfilm). Documents are printed to microfilm, appearing exactly as if they had been printed to paper and then microphotographed. Storage densities are high, storage costs are low, and in the worst case, the images can be read with a source of light and a simple lens.

Information security demands that we be able to read old data: it is time for us to pay serious attention to long-term storage technologies.

[The original version of this article appeared in an issue of the British *Secure Computing* magazine in 1995 and was later reprinted in the medical informatics magazine *For the Record*.]