

WRITTEN ON THE WIND

THANKS TO THE WILD ACCELERATION OF TECHNOLOGICAL CHANGE, DIGITAL INFORMATION IS BEING RENDERED IRRETRIEVABLE ALMOST AS SOON AS IT IS STORED. WELCOME TO WHAT COULD BE THE DAWN OF THE DIGITAL DARK AGE

BY STEWART BRAND

THE PROMISE HAS BEEN MADE:

“Digital information is forever. It doesn’t deteriorate and requires little in the way of material media.” So said one of the chieftains of the emerging digital age, computer-chip maker Andy Grove, the head of the Intel Corporation. Another chieftain, Librarian of Congress James H. Billington, has set about digitizing the world’s largest library so that its contents can become accessible by anyone, from anywhere, forever.

But a shadow has fallen. “It is only slightly facetious,” wrote RAND researcher Jeff Rothenberg in *Scientific American*, “to say that digital information lasts forever—or five years, whichever comes first.”

Digitized media do have some attributes of immortality. They possess great clarity, great universality, great reliability and great economy—digital storage is already so compact and cheap it is essentially free. Many people have found themselves surprised and embarrassed by the reemergence of perfectly preserved e-mail or online newsgroup comments they wrote nonchalantly years ago and forgot about.

Yet those same people discover that they cannot revisit their own word-processor files or computerized financial records from ten years before. It turns out that what was so carefully stored was written with a now-obsolete application, in a now-obsolete operating system, on a long-vanished make of computer, using a now-antique storage medium (where do you find a drive for a 5 1/4-inch floppy disk?).

Fixing digital discontinuity sounds like exactly the kind of problem that fast-moving computer technology should be able to solve. But fast-moving computer technology is the problem: By constantly accelerating its own capabilities (making faster, cheaper, sharper tools that make ever faster, cheaper, sharper tools), the technology is just as constantly self-obsolcescing. The great creator becomes the great eraser.

Behind every hot new working computer is a trail of bodies of extinct computers, extinct storage media, extinct applications, extinct files. Science fiction writer Bruce Sterling refers to our time as “the Golden Age of dead media, most of them with the working lifespan of a pack of Twinkies.” On the Internet, Sterling is amassing a roll call of their once-honored personal computer names: Altair, Amiga, Amstrad, Apples I, II and III, Apple Lisa, Apricot, Atari, AT&T, Commodore, CompuPro, Cromemco, Epson, Franklin, Grid, IBM PCjr, IBM XT, Kaypro, Morrow, NEC PC-8081, NorthStar, Osborne, Sinclair, Tandy, Wang, Xerox Star, Yamaha CX5M. Buried with them are whole clans of programming languages, operating systems, storage formats, and countless rotting applications in an infinite variety of mutually incompatible versions. Everything written on them was written on the wind, leaving not a trace.

Computer scientist Danny Hillis notes that we have good raw data from previous ages written on clay, on stone, on parchment and paper, but from the 1950s to the present, recorded information increasingly disappears into a digital gap. Historians will consider this a dark age. Science historians can read Galileo’s technical correspondence from the 1590s but not Marvin Minsky’s from the 1960s.

It’s not just that file formats quickly become obsolete; the physical media themselves are short-lived. Magnetic media, such as disks and tape, lose their integrity in 5 to 10 years. Optically etched media, such as CD-ROMs, if used only once, last only 5 to 15 years before they degrade. And digital files do not degrade gracefully like analog audio tapes. When they fail, they fail utterly.

Beyond the evanescence of data formats and digital storage media

and spurious “improvements,” and can lose the equivalent of volume of Aristotle. But the practice also builds the bridge between human language eras—from Greek to Latin, to English, to whatever’s next.

Archivist Howard Besser points out that digital artifacts are increasingly complex to revive. First there is the viewing problem—a book displays itself, but the contents of a CD-ROM are invisible until opened on something. Then there’s the scrambling problem—the innumerable ways that files are compressed and, increasingly, encrypted. There are interrelationship problems—hypertext or website links that were active in the original, now dead ends. And translation problems occur in the way different media behave—just as a photograph of a painting is not the same experience as the painting looking through a screen is not the same as experiencing an immersion medium, watching a game is not the same as playing it.

Gradually a set of best practices is emerging for ensuring digital continuity: Use the most common file formats, avoid compression where possible, keep a log of changes to a file, employ standard metadata, make multiple copies and so forth.

Another approach is through core standards, like the DNA code in genes or written Chinese in Asia, readable through epochs while everything changes around and through them. The platform-independent programming language called Java boasts the motto “Write Once, Run Anywhere.” One of Java’s creators, Bill Joy, asserts that the language “is so well specified that if you write a simple version of Java in Java, it becomes a Rosetta Stone. Aliens, or a sufficiently smart human, could eventually figure it out because it’s an implementation of itself.” We’ll see.

THE GOLDEN AGE

BEHIND EVERY HOT NEW WORKING COMPUTER IS A TRAIL OF BODIES OF EXTINCT COMPUTERS, STORAGE MEDIA, APPLICATIONS AND FILES OF DEAD MEDIA

lies a deeper problem. Computer systems of large scale are at the core of driving corporations, public institutions, and indeed whole sectors of the economy. Over time, these gargantuan systems become dauntingly complex and unknowable, as new features are added, old bugs are worked around with layers of “patches,” generations of programmers add new programming tools and styles, and portions of the system are repurposed to take on novel functions. With both respect and loathing, computer professionals call these monsters “legacy systems.” Teasing a new function out of a legacy system is not done by command, but by conducting cautious alchemic experiments that, with luck, converge toward the desired outcome.

And the larger fear looms: We are in the process of building one vast global computer, which could easily become The Legacy System from Hell that holds civilization hostage—the system doesn’t really work; it can’t be fixed; no one understands it; no one is in charge of it; it can’t be lived without; and it gets worse every year.

TODAY’S BLEEDING-EDGE TECHNOLOGY IS TOMORROW’S broken legacy system. Commercial software is almost always written in enormous haste, at ever-accelerating market velocity; it can foresee an “upgrade path” to next year’s version, but decades are outside its scope. And societies live by decades, civilizations by centuries.

Digital archivists thus join an ancient lineage of copyists and translators. The process, now as always, can introduce copying errors

Exercise is always the best preserver. Major religious works are impressively persistent because each age copies, analyzes and uses them. The books live and are kept contemporary by frequent use.

Since digital artifacts are quickly outnumbering all possible human users, Jaron Lanier recommends employing artificial intelligences to keep the artifacts exercised through centuries of forced contemporaneity. Still, even robot users might break continuity. Most reliable of all would be a two-path strategy: To keep a digital artifact perpetually accessible, record the current version of it on a physically permanent medium, such as silicon disks microetched by Norsam Technologies in New Mexico, then go ahead and let users, robot or human, migrate the artifact through generations of versions and platforms, pausing from time to time to record the new manifestation on a Norsam disk. One path is slow, periodic and conservative; the other, fast, constant and adaptive. When the chain of use is eventually broken, it leaves a permanent record of the chain until then, so the artifact can be revived to begin the chain anew.

How can we invest in a future we know is structurally incapable of keeping faith with its past? The digital industries must shift from being the main source of society’s ever-shortening attention span to becoming a reliable guarantor of long-term perspective. We’ll know that shift has happened when programmers begin to anticipate the Year 10,000 Problem, and assign five digits instead of four to year dates. “01998” they’ll write, at first frivolously, then seriously. ■