Information devolution?

Dwain K. Butler, ALION Science and Technology Corporation, Vicksburg, MS

Even though we have phenomenal technology to process, catalog, store, and retrieve information, "we may be living in a period less rich for permanent information than in the 19th century."

—Paraphrased and quoted from David Allison, Director, Smithsonian Institution Division for Information Technology and Society, USA Today, 22 October 1999.

We face enormous challenges with the phenomenal volume and diversity of information being generated. As the above quote indicates, however, in spite of our burgeoning capabilities to process, store, and retrieve information, we are just as quickly losing capability to maintain information permanence. In the strictest sense, evolution refers to gradual or episodic changes in a species, system, or process with time. Most people assume or will make the quick layperson's assertion that evolution involves changing for the better or improving with time. So do we have a word for something that is changing in a negative sense—that becomes worse or diminished in capability/utility with time? How about using "devolution" for a species, system, or process that is actually diminishing in "quality" or capability with time? I contend that the frenzied rush to a totally *digital* "information culture" is devolution, at least in the thrusts to totally abandon more permanent display, storage, and retrieval media. In fact, it is devolution at a pace unparalleled in recorded history—a recorded history that spans 50 000 years from cave paintings of early man to the present.

The phenomenal cave art (Figure 1) at Cheval, France, has two very important characteristics that convey a message to us at the dawn of the third millennium CE: (1) it has *persisted* for at least 30 000 years and (2) it is *interpretable* or *readable*. Thus the residents of prehistoric Europe have successfully passed information to us over a time span of 30 millennia. The historic period is normally defined to have begun approximately six millennia ago with the invention of writing or written records by the Sumerians. Sumerian cuneiform and Egyptian hieroglyphic writing found on rock stelae (Figure 2) and tablets *persist* and are *readable (interpretable)*. Written records from these early civilizations have passed information to us over a time span of six millennia. This information has been preserved because of an extremely *durable recording media* and a readily *interpretable format*.

Figure 1. *Prehistoric cave art, Cheval, France, ca 30 000 yrs BP.*

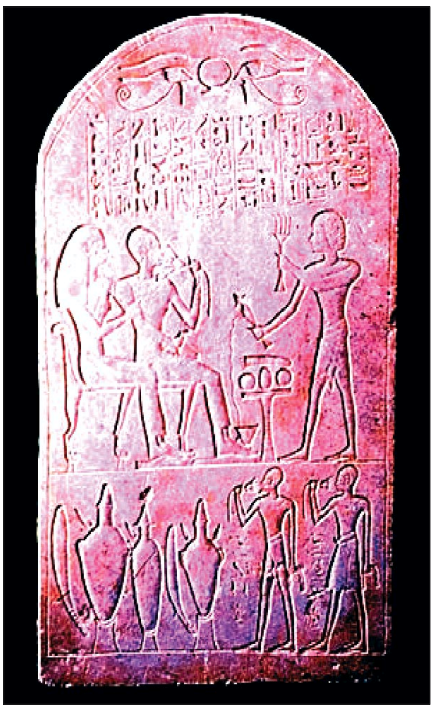

Figure 2. *Egyptian funeral stela, New Kingdom, 18th Dynasty (ca. 1420 BC).*

Information devolution began in earnest with the use of less durable or persistent recording media, organic materials such as bark, animal skins, papyrus, and then paper. The development of paper is generally attributed to Ts'ai Lun, a Chinese official, in 105 CE. By the end of the second century, the Chinese were printing books on rag paper with wooden type. However, until the invention of the moveable metal type printing press by Johann Gutenberg in 1450, bookmaking was still a painfully slow process. *The year 1450 marked the true beginning of the information revolution*, as books and their recorded information were now available to the masses. Even though less durable as a recording media than stone, three-millennia-old scrolls of papyrus persist. And fortunately we have the true wealth of information that exists in book form since 1450. Many centuries-old books are still in remarkable condition and are true works of art, such as the spectacular Gutenberg Bible (Figure 3).

Figure 3. *The Gutenberg Bible, ca 1454.*

To the dawn of the twentieth century, it is easy to trace information devolution from a "shelf life" of > 30 000 years for stone media, to > 3000 years for rolls and scrolls on various organic media, to > 500 years for books on paper. Thus shelf life devolves, but the information is still remarkable for its persistence. With all of the media, the information is encoded or formatted as drawings, pictographs, and letters/words for human visual interpretation, i.e., no intermediary device or machine is needed for interpretation.

The twilight years of the nineteenth century, however, brought two landmark inventions that forever changed the concepts and the potential of information storage and interpretation. With the invention of the phonograph in 1877 and magnetic recording in 1888, the human voice and music could be recorded and played back at will. Originally the phonograph media were wax cylinders, but "plastic" disks (records) replaced the wax media in 1909. The wax cylinders and plastic records have shelf lives > 100 years, since a few of the original wax cylinders and many plastic records (of all playing speeds and genres) still exist. However, the fragility of the phonograph recording media is well

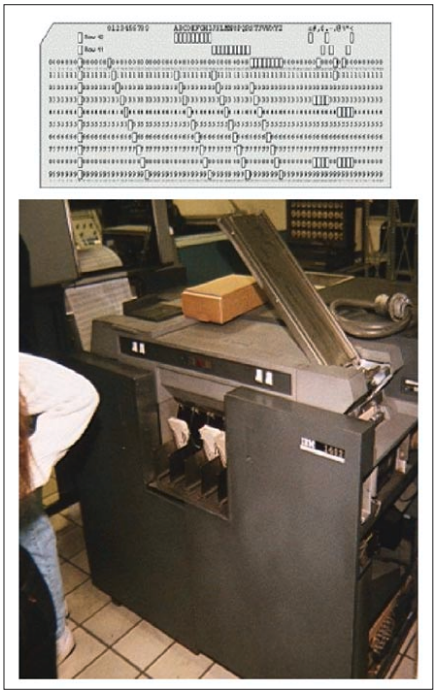

Figure 4. 80-column format punched card and card reader.

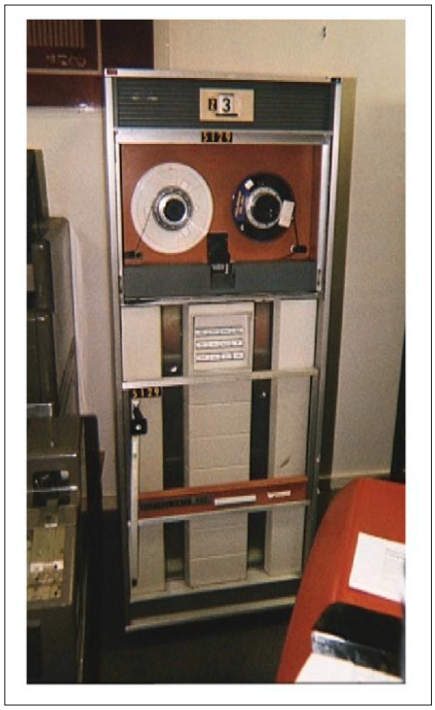

Figure 5. Computer tape drive.

known—e.g., scratching, warping due to heat, breakage. Also, the records require an intermediary device, the phonograph, to interpret the recorded information. *Without* the intermediary device, the recorded "information" may still exist, but it is not interpretable. Fortunately for the nostalgic lovers of the phonograph record, many original phonographs are still functional and new technology phonographs are being produced, many of which replicate the originals in appearance, so the classic 78, 45, and 33 rpm records can still be enjoyed.

The second invention of the late nineteenth century was the Hollerith tabulating machine that was invented to cope with the enormous volume of data involved with the 1890 census. The Hollerith machine could read and tabulate information coded on hand-punched cards, with each card containing the census information of individuals and families. Hollerith's company eventually became IBM in 1924, and the punched card eventually evolved into the 80-column/12-row (don't fold, spindle, or mutilate) IBM standard card format (Figure 4). The encoding changed over time from essentially analog format for the original Hollerith punched cards to binary representation and more sophisticated encoding schemes such as ASCII. For those people who utilized computers during the 1950s, 1960s, and 1970s, card punch machines, card readers, and stacks of punched cards were constant companions. The sturdy punched cards should have had a shelf life equivalent to that of books, but they were made obsolete by rapidly advancing technology. Recently I threw away many boxes of punched cards containing carefully encoded data from earthquake strong motion records. I realized that the cards no longer represented information or even data, since any *practical* means to interpret the encoded data no longer exists except in museums.

In the twentieth century, magnetic recording blossomed, proceeding from the early wound-wire and solid metal cylinders and drums to magnetic tape recording (reel-to-reel and several generations and varieties of cassettes), to computer hard disks, to VHS tapes and other video formats, and to the ubiquitous "floppy" disks. The first magnetic tape drive for a computer appeared in 1951, while the first computer hard drive was developed in 1957. In the late 1970s and early 1980s, with the development of the first PCs, "floppy" disks emerged. Magnetic storage media for voice and music and for computers was and is versatile. Like the phonograph and the punched card readers, however, the device that reads the recorded "information" must be able to accommodate the media, i.e., read it or play it, including interpreting the format of the recording (Figure 5). Practical shelf life of digitally encoded magnetic media is nominally 5-10 years, with the upper limit as the accepted archival life span. Everyone knows that audio tapes have longer life spans, 40-50 years, albeit the quality of playback degrades with time. The frailties of magnetic media are well known: stretching, breakage, scratching, water damage, exposure to stray magnetic fields, etc. But the issue of magnetic media shelf life, like the punched card, begs the question of whether the capability of reading (interpreting) the media will be around in 5, 10, or 50 years. The disappearance of reel-to-reel recorders, eight-track tape players, and large format floppy disk drives has occurred within many of our lifetimes. We are currently witnessing the disappearance of VHS players.

Many of us have experienced the frustration of realizing that a program or data set carefully stored on a 5 1/4 in. floppy disk is no longer accessible; we even held on to an outmoded computer for years after its replacement just for the 5 1/4 in. drive, but finally gave up and let it go. Pathetic is the story of the author (in *USA Today* ca. 2000) who decides to resurrect a manuscript of a novel that was originally typed on a word processor and stored on a floppy disk in the mid-1980s. In the late 1990s when the author decides to publish the novel, he realizes that the word processor has seen several generations of replacements and the disk is no longer interpretable. Fortunately, after much agonizing, the author finds a hard copy of the manuscript, i.e., the manuscript on a recording media with much longer shelf life, and is able to proceed with the publication.

Currently, the recording media of choice are versatile, high capacity hard drives, CDs, and DVDs. As previously mentioned, the hard drives have a practical life span of 5-10 years, while CDs and DVDs and other optical media are estimated to have an archival lifetime of 30 years, although this is certainly an untested estimate. Due to the convenience and high storage capacities of CDs and DVDs, they are the current archival media of choice. Frailties of CDs are familiar. But the far more important issue is the persistence of CD readers and the CD format itself. No one is naïve enough to expect that a CD recorded today will be readable on any computer manufactured 30 years from now (or even 10 years). CDs are so inexpensive to produce and distribute that hard copy publication of proceedings and other types of publications are becoming a thing of the past. Witness the recent arrival of the 2003 *SEG Yearbook* and membership directory on a little plastic disc! (I will carefully save, preserve, and treasure my hard copy of the 2002 *SEG Yearbook*).

The key issue is not one of nostalgia, although for many of us nostalgia certainly plays a role. Who doesn't treasure their book and journal collections? Who doesn't experience a certain feeling of satisfaction and accomplishment and pride at seeing their name in print in a book or journal article? Who doesn't remember their *first* published article? But nostalgia aside, the key issue is data and information persistence (shelf life). We have all heard the estimates on the doubling time of information and knowledge; unfortunately, as indicated in the quote from Allison (1999), we are also *losing* information at an ever-increasing rate. Within each generation of recording media, we are faced with the necessity of "backing up" our stored material in the event of "hard disk crashes" or other media failures and with the questions of which backup media and how often. Also, with each new generation of recording media, we are faced with increasingly massive data/information transfer or migration challenges. *Inevitably, information is lost in both the failure to execute timely and complete backup and the failure to completely migrate. Nowhere is this problem more appreciated and important than in the great information repositories of the world, such as the Library of Congress (Allison 1999; NRC 2000).*

Loss or degradation of shelf life is also evident in the blind rush to online digital publication rather than printed media. Clearly, we have the individual option of downloading and printing, but this does not have the same reliable permanence as published, hard copy archival storage in private, corporate, and public libraries. Also the problem of the proliferation of nonpeer-reviewed, Web-based publishing is growing, and unfortunately it is not always easy to distinguish reviewed versus nonreviewed sources. All professional societies, publishing companies, libraries, and researchers are facing this dilemma. Digital publication is inexpensive and fast, but we cannot allow time and cost to be the dominant drivers. The transition must be carefully designed, documented, and always flexible to allow for alternatives. I recently went through a painful experience with AGU's *Geophysical Research Letters (GRL)*. After submitting the manuscript through the *GRL* online submission procedure, the paper was "rejected" without review due to length concerns. After revising the paper and ensuring that it was well within the four-page printed format requirement, I tried to resubmit; only to find that inexplicably I was locked out of the system either for resubmission or for a new submittal. In frustration I sent the paper as an email attachment, which was rejected because of the new policy of only accepting papers submitted through the online submittal process. The *GRL* editors had not built in an alternative submission procedure, and asserted to me that no one else seemed to be having a problem. About six months later, after submitting the paper to another journal, I noted a short editorial in *EOS* discussing the difficulties encountered by many authors due to "bugs" in the online submittal process and apologizing for the inconvenience. While the SEG experience with online submittal and review has also had "minor glitches" (Gérard Herman's President's Page, March 2003), it has gone smoothly as a result of careful planning, *and there is still an option to submit a hard copy* (although there is apparently a charge for submitting the old-fashioned way!).

In the final analysis, I must confess, if it is not entirely obvious by the tone of this article, that for me the problem is one of deep-rooted nostalgia. It is also definitely a generational issue, as evidenced by a recent comment by a young geophysicist friend—something to the effect, "I'm not interested in books or hard copies, give me a PDF file!" Well, for me, I prefer a hard copy that I don't have to print out on poor-quality copy paper, that I can mark up as desired, that I can use to quickly flip forward or backward without losing my place, and that I can have access to anytime without booting up my computer and trying to remember where the file is located or where I stored the little plastic disc! Also, the wonderful experiences of browsing the shelves in the quiet of a library are deeply rooted in my psyche.

As I review what I have said above, I fully appreciate the capabilities at my disposal for assembling the manuscript of this paper. The capability to perform spelling and grammar checks, easy reformatting of the whole paper, cutting and pasting, inserting figures and text boxes, resizing images, etc. Then after all is complete, I just send the manuscript as an email attachment! So, nostalgia aside, isn't life grand and "modern" information technology wonderful!

> No stereotype of libraries as quiet, uneventful places could survive the 1990s. Whatever stability and predictability libraries once had as ordered storehouses of the treasures of the printed word were shattered by the digital revolution. The intellectual function of libraries—to acquire, arrange, and make accessible the creative work of humankind—is being transformed by the explosion in the production and dissemination of information in digital form, especially over global networks.
>
> —National Research Council, 2000

Suggested reading. "Information age losing memory: Items on outdated systems might be impossible to retrieve," Money Section, *USA Today*, 22 October 1999. LC21: A Digital Strategy for the Library of Congress. Committee on an Information Technology Strategy for the Library of Congress (National Research Council, 2000). TLE

Corresponding author: Dwain.K.Butler@erdc.usace.army.mil

AUGUST 2003 THE LEADING EDGE 0000