

# DIGITAL DATA ARCHIVES...

## ETCHED IN STONE?

Recently, I took my first big road trip in years. As I got into my car armed with a couple of dozen music CDs to pass the time, I recalled my last cross-country trip from San Francisco to Houston in 1980. On that trip, I carried with me dozens of audio cassette tapes that I had diligently recorded from vinyl albums. Now those tapes are at home gathering dust. It may not be too long before I switch to an iPod or other digital music library. Must I migrate everything to a new format to be able to play it in the future?

Considering the exponential growth of environment, health, and safety (EH&S) data, compliance requirements, records retention policies, and legal issues, how can organizations archive and protect EH&S data for retrieval 30, 50, or even 100 years from now?

### DIGITAL ARCHIVES

Times were simpler when most of our data were stored on paper—or were they? Hard copy archives can be retrieved and read without the need for special tools, but it can be tricky to locate the exact piece of paper with the information you need, especially if you store large amounts of data in one place. Also, paper degrades when exposed to heat, light, humidity, and other environmental conditions.

The word “archive” brings to my mind a damp, dimly-lit room stacked to the rafters with musty, yellowing papers—a place to keep historical records. In information technology (IT) speak, archive means files that are no longer in current use, but kept in long-term storage in case they are needed in the future. A digital archive contains data older than the three to five weeks worth of data typically stored in backup and recovery systems.

Common archive media include magnetic tapes, disc drives, and optical media. Because digital archives store data as zeros and ones, they require specific hardware and software to read the data.

### TAPE

IT organizations have used magnetic tapes for decades. Tapes are relatively inexpensive, but are not a state-of-the-art solution for archiving. Data recovery from tape becomes more difficult as the data volume grows. After three to five days, most organizations move tapes to offsite storage in vaults or even deep within salt mines. The shortcomings of archiving

on tape become apparent when someone needs to access a file that has been in storage for some time.

Magnetic tapes are available in many different formats. My colleague, who manages the tape operations for a global company, uses four different mainframe tapes and at least six types of server tapes. The company prefers to use tapes that can back up 400 GB of data in 80 seconds. They load some tapes manually—which is labor-intensive—and use autoloaders for others.

IBM physicist and storage expert Kurt Gerecke claims that magnetic tapes are superior to burnable CDs and that they can last from 30 to 100 years ([http://www.infoworld.com/products/print\\_friendly.jsp?link=/article/06/01/10/73755\\_HNcdlifespan\\_1.html](http://www.infoworld.com/products/print_friendly.jsp?link=/article/06/01/10/73755_HNcdlifespan_1.html)).

### HARD DISCS

Many organizations have opted for hard discs (i.e., servers) for data archiving, since hard disc storage has become more affordable in recent years. Such over-the-network systems allow instantaneous data retrieval and require less IT staff and resources than tape systems. Files on the network are easier to track, faster to index, and easier to transfer than files stored on tape. Large organizations may use an array of centralized servers or have each department or division host its own archive server. Small businesses often maintain archives on small servers or external hard drives with 100 MB or higher capacity. To read data well into the future, keep in mind that the drive bearing (like a ball bearing) limits its life span, so be sure to select a drive that operates at 7200 rpm.

Some companies like EMC Corp. ([www.emc.com](http://www.emc.com)) specialize in hardware/software solutions and help organizations automate the archiving process. Other companies provide online data management, a simple and cost-effective way to provide long-term access to digital records. According to LiveVault ([www.livevault.com](http://www.livevault.com)), a provider of online backup and recovery services, small and medium-size businesses with remote offices are moving to online systems to displace older, manual, tape-based systems.

### OPTICAL DISCS

Organizations, large and small, rely on writable optical discs such as compact discs (CDs) and digital versatile discs (DVDs) for archiving, believing that data are etched in stone. Various vendors claim that their optical discs last anywhere from three to 300 years, but there is little data to back this up

([www.itl.nist.gov/div895/gipwog/Feb-2-05/Imation\\_Edwards.pdf](http://www.itl.nist.gov/div895/gipwog/Feb-2-05/Imation_Edwards.pdf)). Some manufacturers fabricate discs on their own production lines, some contract with others to produce the discs, and others simply purchase discs for resale—buyer beware!

Writable CDs and DVDs (CD-Rs and CD-RWs) differ from original “pressed” discs like music or software CDs. Gerecke says, “Unlike pressed original CDs, burned CDs have a relatively short life span of between two and five years, depending on the quality of the CD.” The recording surface of CD-Rs and CD-RWs has a dye layer that is modified by heat to store data. Dye degradation can result in the data “shifting” on the surface, thus becoming unreadable to the laser beam. “Many of the cheap burnable CDs have a life span of around two years.... Some of the better quality discs offer a longer life span, of a maximum of five years.” Ultimately, quality

control processes and the dyes used in disc production determine their life span.

The Government Information Preservation Working Group (GIPWoG) of the National Institute of Standard and Technology, together with The DVD Association (DVDA), is working to establish a long-term, or archival, standard measurement for recordable CD and DVD media. A 2005 survey sponsored by the GIPWoG showed that consumers would be willing to pay more for high-quality discs that would better preserve their data ([www.dvda.org/html/gipwog.html](http://www.dvda.org/html/gipwog.html)). The GIPWoG proposes to establish industry testing and accreditation standards, plus consumer-friendly labeling that certifies discs and indicates their life span. Standards will differentiate discs according to quality, not price, with high-quality discs labeled as, say, “Archival Grade” or “Storage Quality.”

### NOTHING IS ETCHED IN STONE

Archiving is mission-critical to government agencies, academia, and businesses of all shapes and sizes. Since no digital storage medium lasts forever, consider data volume and growth, compliance requirements, and legal issues when selecting archive media. Seek guidance on IT, archiving processes, and EH&S issues and invest in high-quality media, hardware, and software. Finally, implement a migration plan as new technologies emerge, so that you can read archived data well into the future. **em**



Jill Barson Gilbert, QEP, is president of Lexicon Systems, LLC. She helps organizations increase business value by designing and implementing EH&S management solutions that leverage technology. E-mail: [JBGilbert@Lexicon-Systems.com](mailto:JBGilbert@Lexicon-Systems.com).