Technology-Independent Information Storage

Steven H. McCown
One StorageTek Drive - MS 2129
Louisville, CO 80028- 2129
steven_mccown@storagetek.com

Michael Leonhardt
One StorageTek Drive - MS 2129
Louisville, CO 80028- 2129
michael_leonhardt@storagetek.com

ABSTRACT

The introduction of computer technology into the genealogical domain has facilitated unprecedented advances that have furthered genealogical research to a previously unimaginable level. For the foreseeable future, this computer technology will continue advancing at or surpassing the predictions of Moore's Law (i.e., fundamental performance parameters double every 12-18 months). Database technology will advance, image-processing capabilities will increase, and character recognition will improve to where it is nearly flawless. With all of these advances, the resultant data will be inevitably archived on contemporary computer storage systems.

Common magnetic tape digital storage media is predicted to have an archive life of 20 years or more under controlled environmental conditions. Some optical disk media even claim storage capabilities exceeding 100 years. While this is outstanding for "deep archive" industries, including genealogy, it does present some potentially serious problems. For example, what types of computer systems will be in use 100 years from today, or even 20 years? Will these future systems be capable of reading today's media? In a century, today's state of the art systems will inevitably have been relegated to museums, long forgotten storage facilities and landfills.

John Van Bogart, at the National Archives and Records Administration's 11th Annual Preservation Conference in 1996, gave a speech entitled, " Magnetic Tape Storage" in which he said: *"To ensure that the media will be readable far into the future, it may be necessary to archive the system along with the media. For a 100-year life, recording systems and sufficient spare parts will need to be archived along with the data storage media. Media with life expectancies greater than 20 years are capable of out-surviving existing recording system technologies."* His comments were probably conservative. Just seven years ago many computer systems had 5-1/4" floppy disk drives; it would be difficult to find a contemporary computer to read such media today.

Certainly, preserving a sampling of all digital computing systems will better enable the access of deep archives. However, such a task is expensive, requires continual maintenance, and does not always ensure that the data is retrievable. Degradation of archived systems, environmental hazards, natural disasters and other unforeseen circumstances can all contribute to data loss.

Microfilm has been a historical alternative to paper records and digital media. It has the benefit of up to a 500-year archive life, simple viewer interfaces, and wide acceptance. Obvious drawbacks to microfilm include the cost, the required wet chemical development, batch-only processing, limited storage density, lack of optimization for automation, and a notable difficulty in interfacing with databases.

To assist in the current problems associated with long-term storage, Storage Technology Corporation has developed a proof-of-concept archive system known as the "Century Archive Project" (CAP). CAP provides a combination of traditional digital data storage functionality with some of the beneficial features of analog image storage such as microfilm. The main benefits of CAP include: high density storage of human readable document images and their indices, storage of digitally encoded documents, metadata ascription to aid retrieval, and an automation compatible industry standard physical media form factor.

The CAP process involves initially scanning a source document (which could include black and write, gray scale, color or stereoscopic images) to an image file. In the case of writing digital information, such as a database record, a human readable representation (e.g., ASCII characters) is created. Once in memory, the source document is written, in human-readable form, to an optical Write Once Read Mostly (WORM) tape using a scanning laser. Writing data in this manner allows it to be retrieved using an automated imaging system or by other analog methods as simple as a microscope. This human-readable technology ensures that documents will be retrievable even if current digital technology is unavailable. Collateral benefits of this human-readable technology assist the legal industry where legal documents are still archived as individual sheets of paper due to courts' requirements for actual signatures and not digitally stored equivalents.

Documents and images along with their related indices and metadata can co-exist in both digital and human readable form on the CAP media. While the technologies are available, digitally encoded records provide easier accessibility to digital processing systems and databases. Once digital access means are obsolete, analog techniques would be used.

Analog image storage is normally done by writing a gray scale representation using half-toning techniques. Color documents are stored as separate red, green, and blue images; three-beam optics are required for direct color viewing. Optics are also used to provide direct image viewing of stereoscopic images.

CAP provides an alternative storage method for valued documents and images. This storage method allows for direct optical viewing, eliminates drive, media, and technology migration, and offers robust media options for relaxed environmental storage conditions. The digital storage aspect of CAP ensures faster data availability and easier integration with current digital systems. Since the form factor of CAP media is that of contemporary tape media, it is also compatible with existing robotic tape libraries.