Digital Data Preservation: The Millennium CD and Graceful Degradation

Barry M. Lunt, Ryan Sydenham, Feng Zhang, Matthew R. Linford; Brigham Young University

1. Abstract

Digital information is presently stored on optical disks, magnetic disks, and solid-state memory chips. The expected lifetime for these media is not good, when considering them for archival purposes: recordable optical disks: 7-15 years; magnetic disks: 30-50 years; solid-state memory: 10-12 years. Each of these numbers are for media stored under controlled conditions; the numbers get much worse if the media are stored under ordinary use conditions, which includes being transported and handled.

More and more information is being converted to digital format, or originates digitally. A prime example is digital photos or videos. In the commercial sector, this problem is resolved with migration and backups. However, these practices are rarely followed in the private sector. If the problem of digital data preservation for the masses is not resolved, it will not be possible to pass pictures down to succeeding generations.

Another part of the problem is the lack of graceful degradation for digital media. Either a digital file reads perfectly, or it cannot be read. This means that, as a digital file degrades, its degradation is imperceptible until it is too late.

This paper will report on an interdisciplinary research project at BYU which will create a digital storage medium that will last approximately 1,000 years. The graceful degradation will be useful in all digital storage media, whether archival quality or standard quality, and will allow the user of the digital media to know the status of the data and when it should be copied to new media, as well as recover the majority of the information when the digital media fails.

2. Introduction

A great deal of genealogical and family history research depends on artifacts produced by the people themselves. Examples of these artifacts include personal letters and journals, photographs, and other family records. Historically, these have been created and preserved on paper, which has the advantages of being widely available, inexpensive, and a reasonably long shelf life (generally well over 100 years). The major concern of this paper is the fact that many such documents are presently only created digitally, and never are converted to paper. Examples of these documents include email, digital photographs, computer journals, and blogs.

Because these digitally-created documents are rapidly replacing paper as a primary method of creating personal and family history documents, it is of great interest to the family history community that there be a paper equivalent for digital data: widely available, inexpensive, and a long shelf life. The problem is that there presently is no such medium available; the purpose of this paper is to provide preliminary results on research to solve this problem.

3. Current Digital Data Storage Options

Presently, there are three basic storage technologies available for digital data: magnetic, optical, and solid state. The characteristics of these media need to be understood in order to grasp the seriousness of the present situation.

Magnetic media include hard disk drives and tape drives. Hard disk drives are known to be prone to catastrophic failure; if there is a "head crash", the data on the disk may be forever irrecoverable. However, even if there is no "head crash", the digital data degrades with time.

A little background is necessary here. The preservation of digital data depends on being able to preserve the 1s and 0s of the digital nature of the data. These 1s and 0s are stored as contrasting magnetic states, or optical states, or electron storage states, in the three media types, respectively. As the contrast between these states degrades, which it inevitably does with time, the difference between a 1 and a 0 becomes less and less distinct. When this condition degrades to the point that a 1 cannot reliably be distinguished from a 0, the data becomes forever lost.

The laws of physics, most particularly the second law of thermodynamics (also known as the law of entropy), guarantee that this degradation will occur over time. With magnetic media, the most long-lasting of the three, this process takes between 30 and 50 years (Navale). With recordable optical media, the projected lifetimes are between 7 and 25 years (Byers; Shahani, Manns & Youket; Slattery, Lu, Zheng, Byers & Tang;). With solid state storage, the projected lifetime is about 10 years (Flash Memory). For passing things along to succeeding generations, none of these media is adequate.

Many organizations, aware of this problem, have resorted to multiple copies of the data in multiple places, a strategy very useful where focused attention can be given to the data. But the world of family history, most notably personal sources, are characterized by a store and ignore mentality. Unfortunately, what this means is that 30 to 50 years from now when you find your grandparents' personal computer files (stored on any current media), the data will be gone, and their emails, their computer journals, and their digital photographs, will be lost forever.

4. Proposed Solutions: Millennium CD and Graceful Degradation

Present research being conducted at BYU is focused on a two-fold approach to solving this problem. The first is the millennium CD; the second is graceful degradation.

Present CDs are characterized by problems with delamination of the layers of materials used for the recording, and with degradation of the optical contrast created when the CD was recorded. However, there are solutions to both of these problems, using existing materials and existing processes. Research has already created a CD with layers that will not delaminate, even in the most extreme environmental conditions. Future research will focus on creating a CD with materials that will not degrade for over 1,000 years – hence the term a millennium CD. After completion of the development of a recordable millennium CD, a recordable millennium DVD will be developed, followed by a higher-density recordable millennium disc. These will all be readable in current optical disc readers.

One other significant problem with digital data storage is that the user never knows the actual condition of the data until it is too late; digital data is either 100% readable and correct, or it is unreadable; this would be termed very ungraceful degradation. This is not inherent to the digital data recording process, the medium, nor digital data in general, and research is also focusing on a process for allowing something very different, known as graceful degradation.

Due to intellectual property considerations, further details on these research projects is not available at this time. However, it should be of interest to know that there are no major breakthroughs needed for this research to succeed – no new materials nor manufacturing processes are needed. This research should produce a prototype millennium CD and CD recorder within about one year.

References

Byers, Fred, "Optical Discs for Archiving", www.osta.org/technology/pdf/whitepapers/OSFA-12-05-04.pdf.

"Flash Memory", (no author listed), www.corsairmemory.com/corsair/products/specs/PRODTEMPflashmemv5.pdf.

Navale, Vivek, "Predicting the Life Expetancy of Modern Tape and Optical Media", *RLG DigiNews*, Vol 9, #4 (Aug 15, 2005).

Shahani, Chandru J., Basil Manns, Michele Youket, "Longevity of CD Media: Research at the Library of Congress", www.loc.gov/preserv/studyofCDlongevity.pdf.

Slattery, Oliver, Richang Lu, Jian Zheng, Fred Byers, Xiao Tang, "Stability Comparison of Recordable Optical Discs – A Study of Error Rates in Harsh Conditions", *Journal of Research of the National Institute of Standards and Technology*, Vol 109, #5 (Sept-Oct 2004).