Permanent Storage for Digital Photos

Barry M. Lunt, Brigham Young University, Provo, UT, USA; Douglas Hansen, Millenniata, Inc., American Fork, UT, USA; Matthew R. Linford, Brigham Young University, Provo, UT, USA.

Abstract

Since the dawn of digital data storage, it has always been the case that digital data is rather ephemeral. Organizations deal with this by regularly migrating to new storage, archiving to tape, storing in the cloud, and other techniques known together as active management. While this seems to have solved the problem for organizations, it does not solve the problem for individuals, for whom active management is untenable.

This paper presents the results of research focused at creating a medium for permanently storing digital photos and any other digital data, and compares the results of this study to all other digital data storage options available today.

Introduction

Less than a decade ago, the lead author of this paper purchased his first digital camera. Only a couple of years later, he had hundreds of personally valuable photos, all stored on his laptop's magnetic hard-disk drive (HDD). Research into permanent digital data storage began shortly after he realized that there did not exist a way to preserve these pictures for generations.

Determining how long something will last has long been a very important area of study for science and technology, particularly in materials and coatings. Many advances have been made, and much is known today about how to reliably predict the life expectancy (LE) of a product, based on the materials used to make it and the conditions of its use. These advances are readily applied to the field of data storage.

Ideal Characteristics for Digital Photo Storage

There are at least six characteristics that are important for personal storage of digital photos: 1) no active management required (store and forget); 2) no special storage conditions required (store it anywhere); 3) no power or money required to maintain data (related to active management); 4) a life expectancy of at least 100 years, preferably 500 (long lifetime); 5) easily transported (light and reasonably durable); 6) likely to be readable for 500 years.

There are many storage solutions today, but only one (the results of this research) is capable of meeting all these requirements.

Causes of Failure

In the end, all digital data is stored as an optical, magnetic, charge, or other physical property of a certain type of material. The most common failure mechanisms for materials (excluding mechanical wear – poor handling) include oxidation, corrosion, and breaking of chemical bonds. Each of these failure mechanisms is exacerbated by elevated temperature, humidity, and exposure to light. And that is the reason that any environment that is intended for archival storage always includes controlled temperature, humidity and light. These same failure mechanisms come into play when we consider how to store digital data.

Predicting Life Expectancies – Magnetic Tape

In his paper, "Predicting the Life Expectancy of Modern Tape and Optical Media"1, author, Vivek Navale looked at multiple studies on LE for magnetic tape[2,3]. According to one of these studies, "Every tape type showed a loss in magnetization when they were under induced stress conditions of higher temperature and relative humidity."[1] Based on the results, the author reported LEs ranging from 10 – 200 years (depending on the type of tape used) if stored at 30°C; this range decreased dramatically to 0.7 – 7 years if stored at 60°C.

As previously stated, these LE calculations were based on the measured decrease in magnetization. Another commonly used measure for LE calculations is the errors before correction, reported on magnetic tape as the Block Error Rate, or BLER. The studies showed a clear increase in the digital errors while these tapes were stored at 40°C/50% RH. This increase clearly means that these tapes would eventually fail to read the data back correctly. The calculated LEs ranged from 9.3 years to 1083 years – a HUGE range.

In summary, the minimum LE for magnetic tape is under 10 years. It is the minimum we care about the most here – the point at which the first data loss occurs.

Predicting Life Expectancies – Hard-Disk Drives

Hard-disk drives (HDDs) store the majority of the digital data in the world today. Unfortunately, most computer users are all too familiar with the fact that HDDs have a nasty tendency to fail catastrophically. But what can be said for their LE?

The best way to predict the LE would be to monitor many HDDs in normal use for a long time, long enough to see their characteristic failure statistics. Just such a study was done a few years ago.

In their 2007 study, Pinheiro et al. reported on a very large population of HDDs in service at Google, Inc. – over 100,000 of them. Figure 1 shows the Annualized Failure Rates (AFR) for these HDDs.

The data reported in their study is insufficient to project an average expected LE for their population of HDDs, but it is obvious from Figure 1 that there is a fairly wide distribution, and that some drives (about 2.5%) fail as early as 3 months. This clearly eliminates HDDs as an archival storage option.

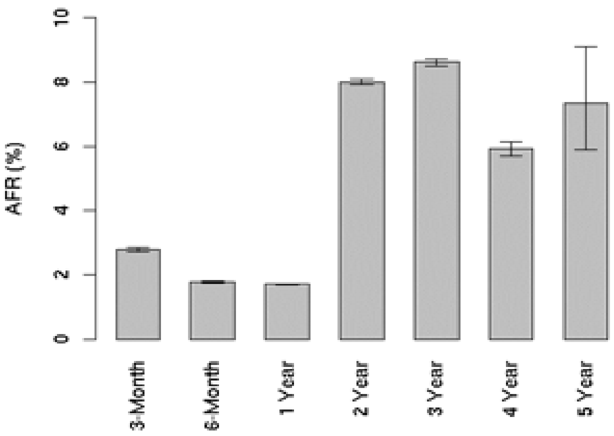

Figure 1. *Annualized failure rates broken down by age groups.*

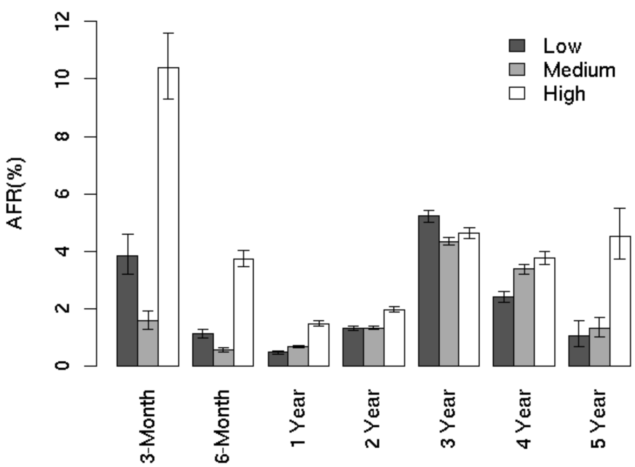

Figure 2. *Utilization AFR for the Pinheiro et al. study.*

Predicting Life Expectancies – Flash Memory

In about 1984, EEPROM was introduced. In those early days, it was well known that due to the fact that the data was stored as the charge on a very small and somewhat leaky capacitor (the floating gate), the Mean Time to Data Loss (MTTDL) for EEPROM was in the range of 10-12 years. Since those early years, many changes have been made, and Flash has become the dominant form of EEPROM. Densities have risen dramatically, from the early 256kb capacities, to today's 8GB capacities and more, all in a single chip. But even in today's Flash memory, the MTTDL has not changed that much, due to the way in which data is stored. For example, in their 2008 article, Kaneko et al.[6] report that the MTTDL for a Flash SSD (solid-state drive) is approximately 13 years.

While the MTTF (Mean Time to Failure) of the actual devices themselves (over 100 years) is much longer than the MTTDL, the issue here is that, without active management, the data on SSDs literally evaporates with time, and that evaporation time is well understood.

We should also mention here that the same is true for Flash memory sticks, also known as jump drives, USB drives, or USB sticks – they all store data for only about 10-12 years, since they all use the same basic floating-gate architecture for storing each bit.

Predicting Life Expectancies – Stamped Optical Discs

Data on stamped optical discs, whether CDs, DVDs, or BDs, is recorded at the time the discs are manufactured, and cannot be altered by the optical disc drives. This type of optical disc is commonly referred to as ROM (Read-Only Memory), since it cannot be recorded by the user.

In his 2005 paper, Navale[1] reported the LE of CD-ROMs to range from 20 to 12,000 years, with a mean LE of 1592 years. While the mean is outstanding, the distribution ranging as low as 20 years is very problematic. Clearly, for archival purposes, research needs to be performed to determine the causes of the early failures. If these causes can be addressed, and the lower end of this distribution fixed, this format of digital data storage could easily be the longest lasting of all current options.

As appealing as it is for a storage medium to have an average LE of over 1,500 years, it is simply not practical for individual users. The reason is that the recording process is the manufacturing process, which means it is very costly for the equipment, and completely impractical for low volumes.

Predicting Life Expectancies – Recordable Optical Discs

There have been several publications that have addressed this area, and with good reason. Recordable optical discs, and the drives to read and record them, are widely available, inexpensive, easily transported, and almost ubiquitous. Billions of these discs are sold every year, in all three densities (CD, DVD and BD). With that many advantages, optical discs are a strong candidate for an archival storage, if only the LE of the data is sufficiently long.

Recordable optical discs use a very different data storage mechanism than

Recordable optical discs use a very different storage mechanism than stamped optical discs. A light-sensitive dye is the recording layer. This dye is normally a poor reflector of light. When the dye is illuminated by the right wavelength of laser light, it becomes much more transparent, thus reflecting better off the reflective layer.

The dye used is necessarily very sensitive to light, as it must respond to the laser light in only a few nanoseconds. While this is great for making practical recordable optical discs, it has some serious archival handling implications. If recordable discs are not stored in dark conditions, this dye will degrade, and the recorded data will begin to fade. This degradation mechanism is commonly referred to as dye fading, and is well known in the optical storage industry.

Research at the National Institute of Standards and Technology[8], published in 2004, looked at a set of seven brands of recordable CDs and DVDs randomly selected from the commercial market. These discs were subjected to conditions of accelerated aging, consisting of either elevated temperature and humidity or full-spectrum light. They monitored the digital error parameters of BLER for CDs, and PIE (PI sum 8) for DVDs. After 500 hours of accelerated aging in elevated temperature and humidity, all brands of CDs had exceeded the BLER limit of 220.

After 1000 hours of accelerated aging in full-spectrum light, all but two brands of CDs had exceeded the same BLER limit. For the three brands of recordable DVD studied, two brands had exceeded the PIE limit of 280 after 250 hours in full-spectrum light; the same two brands exceeded this PIE limit after 125 hours at elevated temperature and humidity. Their basic conclusion was that:

"*Depending on the media type and intensity of the light, a disc may fail due to exposure to direct sunlight in as little as a few weeks. This will be especially true when coupled with the heating effect of exposure to sunlight or combined with any other heat source." (Slattery et al., p. 523)*

In their 2004 paper, Shahani et al.[9] studied CDs randomly selected from a collection of over 60,000 CDs, to determine if the digital errors on these discs were increasing. This was determined by monitoring the Block Error Rate (BLER). They noted that the average BLER had increased from 70.5 (in 1996) to 72.4 (in 1999), and to 74.4 in 2003. While none of these values exceeded the maximum specification of 220, it was a concern that there was a steady upward trend in that number.

Of particular interest in the Shahani et al. study was their characterization and pictures of the failure modes of the discs in their collection. These failure modes were corrosion of the metal layer, oxidation of the reflective layer, and delamination. The discs included in this study were mostly CD-ROMs, so dye fading was not a significant factor. But the other degradation modes have been shown to be definitive for optical discs in general.

Another very significant study on recordable DVDs was released in 2009 by Svrcek of the Naval Air Warfare Center Weapons Division in China Lake, CA[10]. They tested 25 discs from each of six brands of DVDs, including Delkin, MAM-A, Mitsubishi, Verbatim (all archival-quality DVDs), Taiyo-Yuden (a top-rated standard-quality DVD) and Millenniata (advertised to be truly permanent). These 150 DVDs were subjected to accelerated aging conditions of 85°C, 85% RH, and 1120 W/m² of full-spectrum light, all simultaneously. After only 48 hours of such testing, their results were: "All dye-based discs failed according to the ECMA PI8 max limit of 280. The post-test error statistics show all Millenniata discs pass the ECMA standard. The data recorded on these disks was recoverable. The Millenniata disks were the only ones tested that maintained information integrity." Figure 3 shows their summary graph of the PI8 max values, clearly showing the degradation of the data on five of the brands of DVDs tested, while the Millenniata discs were apparently unphased.

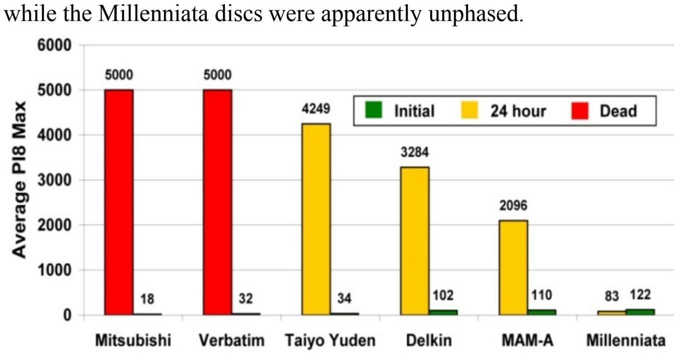

Figure 3. *PI8 Max average by manufacturer including dead discs.*

Based on the preceding research, it is apparent that with one notable exception, recordable CDs and DVDs are currently not in a position to serve as a permanent storage solution for digital data. The one notable exception is Millenniata discs tested in the Svrcek report.

Actual Life Expectancies – Recordable Optical Discs

Accelerated testing to determine LE for optical discs has been used for many years. It is also very insightful to test discs stored in normal conditions to gain another insight into actual LEs. To study this, the authors have been able to obtain data on the recordable optical disc collections of two university libraries. These libraries have chosen to remain anonymous, due to the sensitivity of the information in their possession and their desires to retain it indefinitely.

Library #1 has a collection of approximately 8,500 DVDs, all recorded since 2003. These discs are of eight different brands, including both standard quality and archival quality. These discs are non-circulating discs, meaning that they have constantly been stored under conditions of controlled temperature, humidity and light exposure. These discs were accessed in 2010, and 177 of them had readback problems ranging from unreadable files to unreadable discs. So, in a maximum of 7 years of natural, controlled aging, 2.1% of their collection had experienced permanent data loss.

Library #2 has a collection of approximately 18,000 optical discs, all archival quality, and nearly all of them CDs (about 99.5%). It should be noted that the expected LE for CDs is longer than for DVDs. As with library #1, these discs were non-circulating discs, so they never experienced storage conditions other than controlled temperature, humidity and light. They ranged from 4-14 years old when they were tested in 2008/2009. Of this collection of discs, 319 had either a permanent loss of data or they were completely unreadable. So, in another example of discs kept at conditions of controlled aging, in 4-14 years 1.79% had experienced permanent data loss.

Standard-quality CDs have an LE of approximately 25 years; for archival-quality CDs, the LE is advertised to be 300 years. For standard-quality DVDs, the LE is about 10 years; for archival-quality DVDs, the LE is advertised to be 100 years. Using these values for the average LE for each type of disc, we can estimate a distribution for the LE for each disc type.

First, we assume a normal distribution in the LE for all of the discs in the collections of these libraries. Next we assume the average lifetime of the failed discs from library #1 to be 7 years (which is very generous, considering that some of them surely failed much sooner than that). Then we take the average LE of these discs as a whole to be 100 years. Using these assumptions, we can calculate that the standard deviation would be about 46 years. This means that the consumer has basically no idea when their first data failure will actually occur if their data is stored on these discs.

Doing the same for the CD collection of library #2, assuming the average LE to be 300 years (they were primarily archival-quality CDs), assuming the average LE of the failed discs to be 10 years, and knowing the failure rate to be about 1.79% at 10 years, we can calculate the standard deviation to be about 138 years.

Again, these results indicate that the consumer has basically no idea when to expect the failure of their data. Curators from these institutions have reported that they have observed some discs to fail in as little as two years, which can be expected given the standard deviations calculated above.

Our research leads us to argue that the most important parameter to consider when choosing optical disc media for long-term data storage is not the average LE of the media. The previous section has shown that even CDs with an average LE of 300 years, or DVDs with an average LE of 100 years, can be expected to have a failure rate as high as 2% in only a few years. We emphasize that it is not the average LE that consumers really care about, but that is the only parameter that is advertised or available. The most important parameter to consider when choosing optical disc media for long-term storage is the minimum LE for the discs – but this is never reported, and the actual values are scary. We believe that the only two solutions for making optical discs with a reasonable minimum LE is to both dramatically increase the average LE (to greater than 500 years), and to dramatically reduce the standard deviation. Needed is a storage medium that can store data for a minimum of 100 years[1].

Research on a Permanent Storage Solution

Our research on a permanent storage solution has addressed all the known weaknesses in today's recordable optical discs. The most significant of these weaknesses is the dye, which is organic and light sensitive, and therefore inevitably prone to degradation. We have completely replaced the organic dye with proprietary inorganic material; these materials are known to be stable for extremely long periods of time, and are not light sensitive. Figure 4 shows the layer structure of our recordable DVD-compatible disc, where it can be seen that there is no dye layer, and the reflective layer has been replaced with a 3-layer structure including 2 dielectric layers and a recording layer.

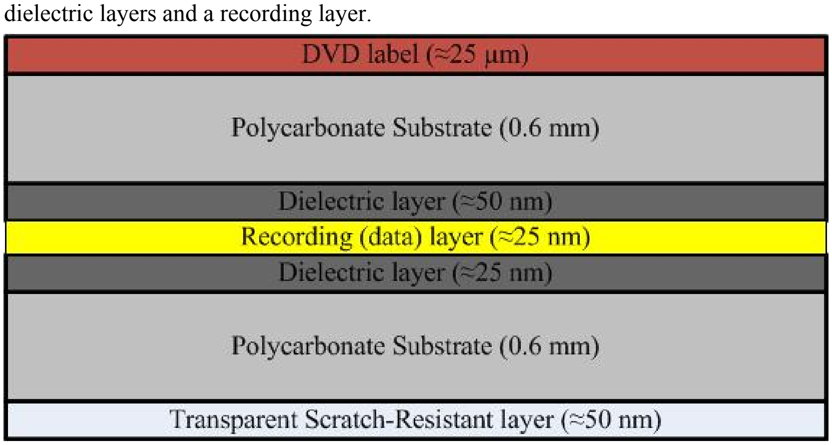

Figure 4. *Layer structure of recordable DVD-compatible Millenniata disc*

This solution also addresses the degradation mechanisms of corrosion and oxidation, as the materials used have been specifically chosen to be extremely resistant to corrosion and oxidation. The scratch-resistant layer addresses the handling problem. In short, this totally new type of recordable optical disc directly addresses all the degradation mechanisms, and has been shown to be dramatically superior to all other digital data storage options as tested by accelerated aging.

One question that may arise is how the Millenniata disc can be recorded, if the recording layer is not sensitive to light. The answer can be seen in Figure 5, which is an SEM of the tracks on the Millenniata disc. The material of the recording layer has been moved by the energy of the laser, creating a permanent absence of material in the locations of the marks, and also creating the necessary optical contrast for an optical disc. Research on drives and write strategies for these discs have been able to bring to jitter to well under the specification of <10%, and the initial number of PI Sum 8 errors to well under 80, giving a degradation headroom of over 200.

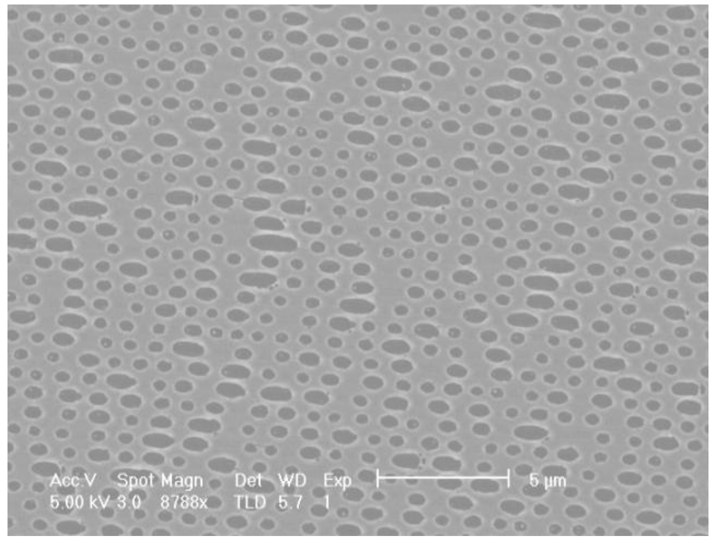

Figure 5. *SEM of a recorded Millenniata disc, showing the physical marks left by the recording process.*

Life Expectancy Summary

The answer to the question that constitutes the title of this paper is found in Table 1. These figures are derived from Navale[1], Van Bogart[2], Pinheiro[5], Slattery[8], Shahani[9], Byers[11], Iraci[12], and Tanaka[13]. The LE values are given as ranges because there are many values reported in these reports. These values are also approximate, for the same reason.

Table 1: Life expectancy for data stored on today's media.

Media	Life Expectancy of Data
Magnetic tape	10-50 years
Magnetic hard-disk drives	1-7 years
Flash drives and Solid-state drives	10-12 years
Recordable optical discs	1-25 years
Millenniata recordable optical discs	1,000 years (advertised)

Format Obsolescence

One other topic that must always be addressed in discussing the archiving of digital data is format obsolescence. There are many examples today of data-storage formats which are now obsolete and are therefore very difficult or impossible to read. Probably the best argument in this area has been made on the blog, "Dr. Barry Lunt's Optical Blog" (http://opticalblog.groups.et.byu.net/), under the posting "Only Half a Solution?" Here Dr. Lunt (one of the authors of this paper) argues that "looking at the past is the best way we have of predicting the future." The past tells us that the most powerful way to guarantee that data will be readable far into the future is found in how widespread the adoption is. A classic example is found in the Latin language. Though no country or people have spoken the language for over 1000 years, scholars are trained to read and speak it today because of how widespread its adoption was, and because there are tens of thousands of documents in existence today which were written in Latin.

In the case of data storage, by far the most widespread formats in the history of digital data are the three main optical disc formats: CDs, DVDs and Blu-ray discs. There are billions of readers in use today, and hundreds of billions of discs; no other storage technology or format even comes close. If the data on these discs persists into the future, history makes a very strong argument that we as a civili zation will not lose the ability to read the data on these discs.

While this very widespread adoption of optical discs as a digital data storage technology does not guarantee persistence far into the future, there is much historical evidence to support the extremely high probability that this format will be readable far into the future.

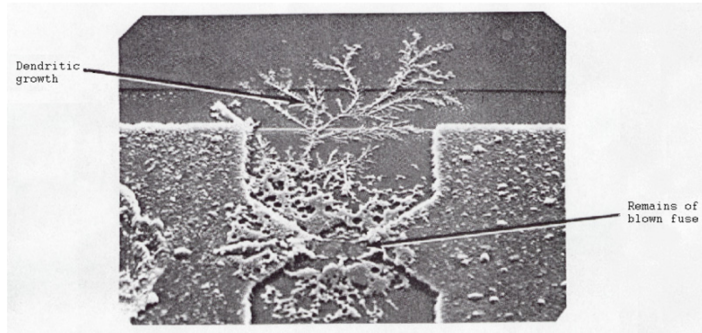

Figure 6. Dendrite growing out of remains of blown fuse. [15]

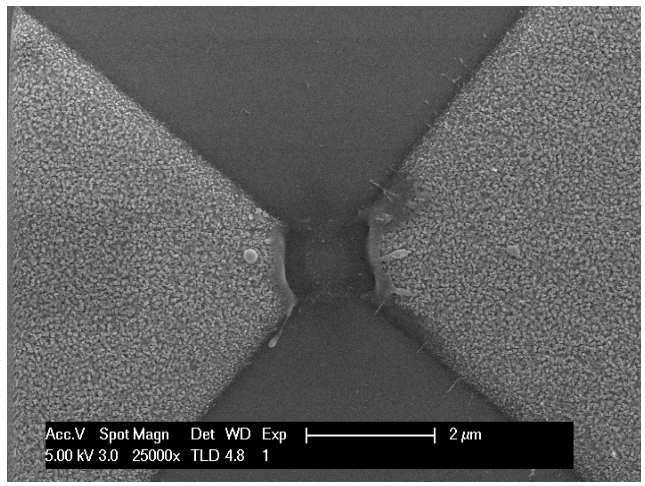

Figure 7. Blown fuse with no material remaining.

Additional Research

Optical discs, while widely adopted, are not the only format for data storage. Flash memory has already been discussed; most research in this area shows that the expected lifetime of the data (if not refreshed) is 10 -12 years.

The alternative proposed is programmable read-only memory (PROM), which has been around for decades, but which has suffered from a failure mechanism known as dendrites (see Figure 6). Dendrites grow in the presence of a voltage field where there is also material present to enable their growth. The present focus is on developing a fuse material that will leave no remnants, as shown in the research SEM in Figure 7. In this example of a blown fuse (a programmed bit location), there are no remaining fuse materials, which eliminates the possibility of a dendrite growing and altering the programmed data.

The future of this permanent solid-state storage looks very good. When development is completed and manufacturing begins, it is estimated that densities and prices will soon be at a par with flash memory. Additionally, due to the ability to make vertical stacks of this memory structure and thus enable 3-D memory, densities and prices would soon be less than those of flash memory.

Summary

Most archivists would prefer to have a storage life of at least 100 years for digital data[14]. According to Table 1, there is only one option for that kind of lifetime, and while the Svrcek report clearly shows it is superior to other recordable optical discs, no LE study has yet been performed on this new technology. One of the biggest problems evident in Table 1 is that users have no way to determine if their media will have an LE on the low end of the distribution, yet that is just as likely as an LE on the high end of the distribution.

References

[1] Navale, Vivek, "Predicting the Life Expectancy of Modern Tape and Optical Media", *RLG DigiNews*, Aug 15, 2005, 9:4; also at www.rlg.org/en/page.php?Page_ID=20744#article3

[2] Van Bogart, John W.C., "Media Stability Studies," National Media Lab Technical Report RE-0017, pp. 1-86, **1994** .

[3] Weiss, R. D. , et.al., "Environmental Stability Study and Life Expectancies of Magnetic Media for Use with IBM 3590 and Quantum Digital Linear Tape Systems," Final Report to the National Archives and Records Administration, Delivered by Arkival Technology Corporation under contract requirement NAMA-01-F-0061, **2002** , pp. 1-94.

[4] Wigmore, Ivy, "How Much Digital Data Is There in the World?", http://itknowledgeexchange.techtarget.com/whatis/how-much-digital-data-is-there-in-the-world-soon-to-pass-the-zettabyte-mark/, May 12, 2010.

[5] Pinheiro, Eduardo, Wolf-Dietrich Weber, Luiz André Barroso, "Failure Trends in a Large Disk Drive Population", *Proceedings of the 5th USENIX Conference on File and Storage Technologies* (FAST '07), Feb 2007.

[6] Kaneko, Haruhiko, Takuya Matsuzaka, Eiji Fujiwara, "Three-Level Error Control Coding for Dependable Solid-State Drives", *Proceedings of the 14th IEEE Pacific Rim International Symposium on Dependable Computing*, Taipei, Taiwan, Dec 15-17, 2008.

[7] Worthington, Mark, "Optical Disc Technology: CD, DVD, BD and Beyond", presentation given to the BYU Harold B. Library, March 2005.

[8] Slattery, Oliver., Richard 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*, 109, 517-524, 2004 .

[9] Shahani, Chandru J., Basil Manns, Michele Youket, "Longevity of CD Media: Research at the Library of Congress", Preservation Research and Testing Division, Washington, DC.

[10] Svrcek, Ivan, "Accelerated Life Cycle Comparison of Millenniata Archival DVD", Life Cycle and Environmental Engineering Branch, Naval Air Warfare Center Weapons Division, China Lake, CA, Nov 10, 2009.

[11] Byers, Fred, "Optical Discs for Archiving", Information Technology Laboratory, Information Access Division, NIST, OSTA, Dec 6, 2004.

[12] Iraci, Joe, "The Relative Stabilities of Optical Disc Formats", *Restaurator: International Journal for thePreservation of Library and Archival Material*, 2005, 26:2, 134-150.

[13] Tanaka, Kunimaro, "Toward Adoption of Optical Disks for Preservation of Digitized Cultural Heritage", *Proceedings of 2008 Optical Data Storage Conference*, Honolulu, HI.

[14] Peterson, Michael, Gary Zasman, Peter Mojica, Jeff Porter (100 Year Archive Task Force), "100 Year Archive Requirements Survey", January 2007, SNIA's Data Management Forum, www.snia-dmf.org/100year, p. 1.

[15] Pecht, Michael, "Handbook of Electronic Package Design", p. 597.

Author Biographies

Barry M. Lunt *has a BS in Electronics and an MS in Manufacturing, both from BYU (Provo, UT). He worked as a design engineer for IBM in Tucson, AZ for seven years, and has been teaching at the college level for 25 years. He has a PhD in Education from Utah State University (Logan, UT). He is presently a full professor of information technology at BYU, where his primary research area is permanent digital data storage. He is also a founder of Millenniata, Inc.*

Douglas Hansen *serves as the Chief Technology Officer and a Founder of Millenniata, Inc. He has over 20 years of experience in High Tech business. He has a BS, MS and PhD in physics.*

Matthew Linford *is an associate professor of chemistry at Brigham Young University. He obtained his B.S. in chemistry from BYU before receiving an M.S. in materials science and a Ph.D. in chemistry, both from Stanford. He is an expert in thin film preparation and characterization. He is also a founder of Millenniata, Inc.*