# How Long Is Long-Term Data Storage?

***Barry M. Lunt; Brigham Young University; Provo, UT, USA***

## Abstract

*In the context of archiving of physical documents, long-term storage has long been accepted to mean centuries. Digital documents are much more ephemeral, so archivists should be aware of the inherent limitations of the technologies available for preservation of digital data. This paper compiles the results of several studies on this subject, in addition to presenting new findings on what can be expected for recordable optical discs (CDs and DVDs). The bottom line is that, with one notable exception, digital data cannot be expected to endure using any existing technologies.*

## Introduction

Less than a decade ago, this author 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). It was then that he asked the question of how to best preserve these pictures for future generations. Being intimately familiar with all the types of storage available, he was immediately distraught to realize that there did not exist a way to preserve these pictures for at least 100 years, preferably much longer. It was also then when he began research into this problem.

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.

## Causes of Failure

The most common failure mechanisms for materials (excluding mechanical wear) include oxidation, corrosion, and breaking of chemical bonds. Each of these failure mechanisms is exacerbated by elevated temperature, humidity, and exposure to light. That is the reason that any controlled 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. There are three basic technologies available for storing digital data: magnetic (including magnetic tape and hard-disk drives), solid-state (consisting primarily of flash memory), and optical (including CDs, DVDs, and Blu-ray discs (BDs)). Each of these technologies uses well-known processes and materials to manufacture the storage media, and each of these technologies has known failure mechanisms, which have been studied.

## 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], which included IBM 3480 and 3590 data cartridges, DLT IV cartridges, SuperDLT cartridges, 8mm data cassettes, D1 and D2 digital video cassettes, and standard VHS videocassettes. According to one of these studies, "Every tape type showed a loss in magnetization when they were under induces stress conditions of higher temperature and relative humidity." [1] Based on their results, they 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.

Navale also showed results from testing these same magnetic tapes at 50°C, with various levels of relative humidity (RH). At RH = 20%, the range was from about 0.6 – 2.8 years; at RH = 80%, this decreased dramatically to a range of about 0.3 – 0.9 years.

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, but Navale pointed out that this prediction is based on the ability of the error-correction coding (ECC) to correct the errors, and is therefore somewhat subjective.

## Predicting Life Expectancies – Hard-Disk Drives

Hard-disk drives (HDDs) store the majority of the digital data in the world today – an estimated 1 Zettabyte (about $10^{21}$ bytes) [4], an amount that is beyond our ability to comprehend. Unfortunately, most computer users are all too familiar with the fact that HDDs have a nasty tendency to fail, and to do so catastrophically. But what can be said for their LE?

The best way to predict the LE would be to monitor many HDDs 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. [5] reported on a very large population of HDDs in service at Google, Inc. – over 100,000 of them. They gathered data on environmental factors (such as temperature), as well as the many parameters that were reported through self-monitoring and analysis software they deployed on their entire system. Figure 1 shows the Annualized Failure Rates (AFR) for these HDDs.

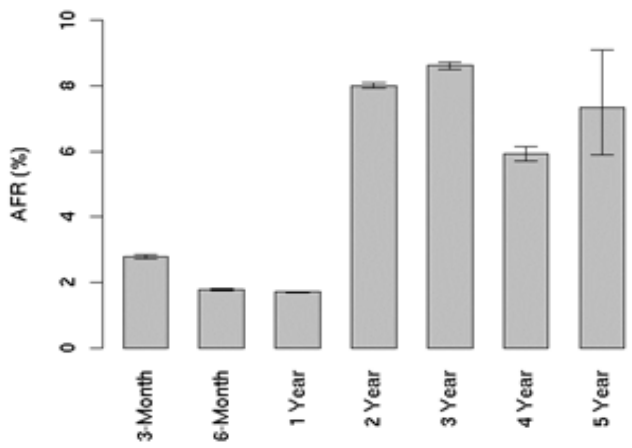


**Figure 1**: *Annualized failure rates broken down by age groups.*

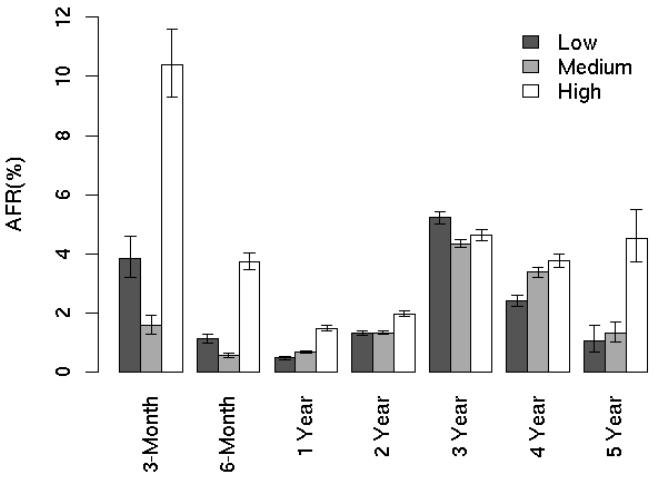


**Figure 2**: *Utilization AFR for the Pinheiro et al. study.*

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.

Another interesting finding in the Pinheiro et al. study is that they found no correlation between utilization and AFR. This is counterintuitive, but Figure 2 shows the AFR as a function of utilization. They categorized utilization into three levels: low corresponds to the lowest 25$^{th}$ percentile; medium corresponds to the 50-75$^{th}$ percentile, and high corresponds to the top 75$^{th}$ percentile. It can be seen in Figure 2 that there is no correlation between utilization and AFR, which means that simply making sure that a HDD is rarely used is not guaranteed to insure it will last longer.

## Predicting Life Expectancies – Flash Memory

In the very early days of non-volatile solid-state memory, EEPROM made a big splash by being byte-level erasable. This was approximately 1984. In those early days, it was fairly well understood that due to the fact that the data was stored as the charge on a very small, 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, all in a single chip. But even in today's Flash memory, the MTTDL has not changed that much, due to the intrinsic 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 is much longer than the MTTDL (over 100 years), the issue here is that, without active management, the data on SSDs literally evaporates with time, and that evaporation time is well understood.

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. Figure 3 give an example of the data structures on a CD-ROM, as seen with a scanning electron microscope (SEM). They are nearly impervious to change, except by extreme conditions.

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.

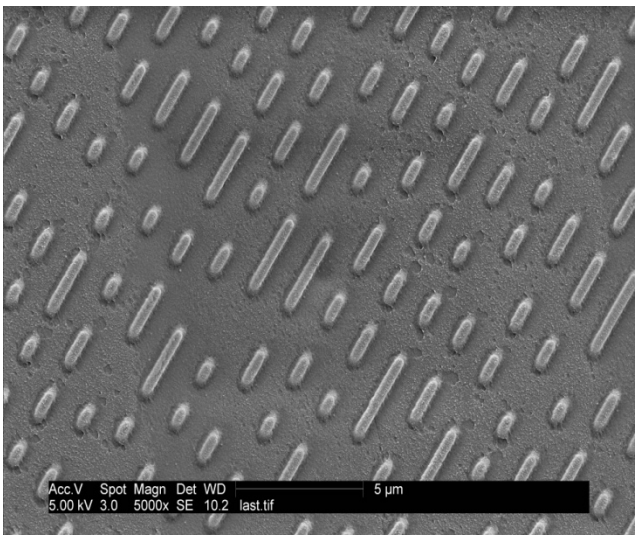


*Figure 3: SEM image of the reflective layer bumps on a commercial CD.*

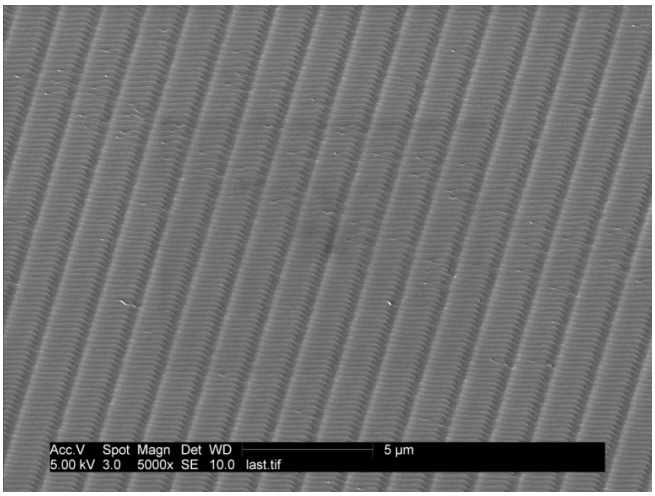


*Figure 4: SEM image of an unrecorded CD-R.*

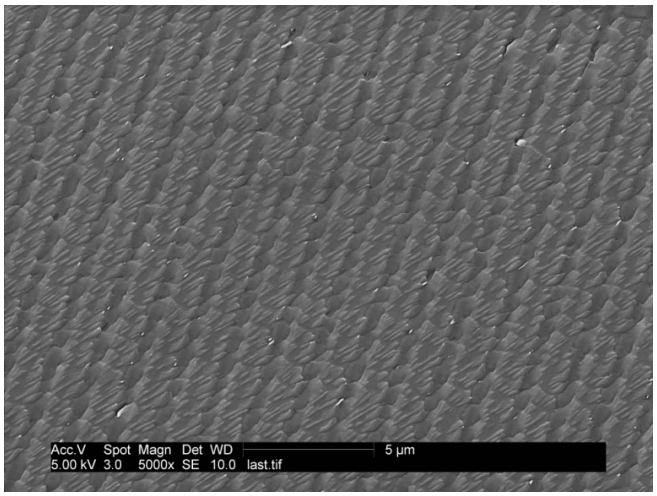


*Figure 5: SEM image of a recorded CD-R*

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 LE of over 1,500 years, it is simply not practical for most 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 them are sold every year, in all three densities (CD, DVD and BD). With that many advantages, they are a strong candidate for archival storage, but only if the LE of the data is sufficiently long. Recordable optical discs use a very different data storage mechanism than stamped optical discs. Figures 4 and 5 show this very effectively as they use the same instrument to image an unrecorded and a recorded CD-R disc. It is obvious that the recording process has physically altered the tracks, but it is also obvious that the data is not to be found in these physical alterations, for there are no discernible patterns.

Optical discs use a layer structure that is very different from ROM discs; this structure is shown in Figure 6. The dye is the optically active component of this structure. The dye is normally a poor reflector of light, as seen in the unrecorded portion of Figure 7. When it is illuminated by the right wavelength of laser light, the dye becomes much more reflective, as seen in the recorded portion of Figure 7.

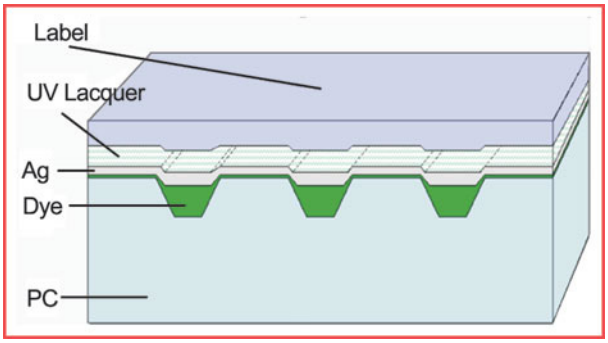


*Figure 6: Layer structure of a CD-R optical disc. (from Worthington)*

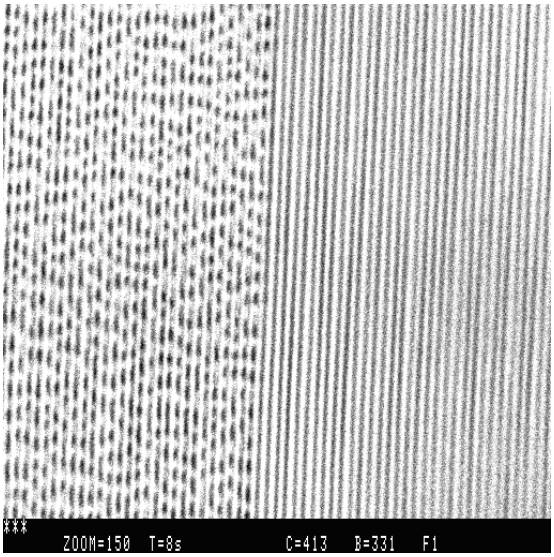


*Figure 7: Unrecorded (right side) and recorded portions of a CD-R as seen through a confocal microscope.*

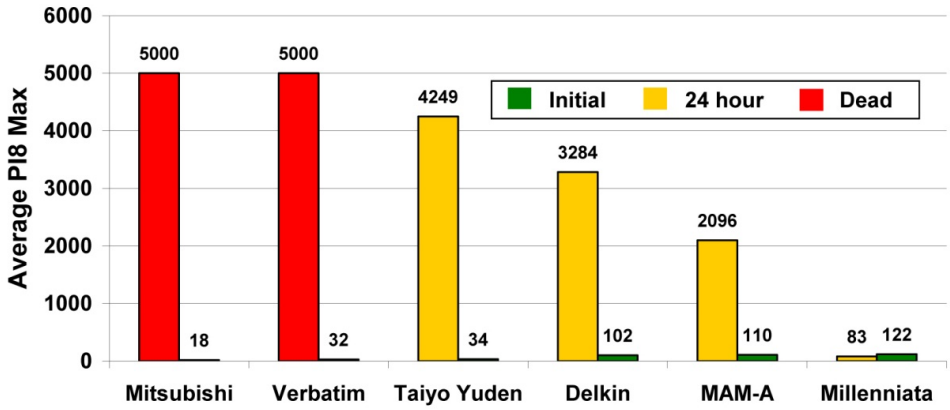


*Figure 8: PI8 Max average by manufacturer including dead discs.*

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 (various combinations of 60°C – 90°C and 70% RH – 90% RH or metal-halide (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 DVDs 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 randomly selected CDs 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 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." (p. 45) Figure 8 shows their summary graph of the PI8 max values, clearly showing the degradation of the data on five of the brands of DVDs tested.

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.

## 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 addressed by the author 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 (also the author 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. And 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 BDs. There are billions of readers in use today, and hundreds of billions of discs; no other storage tech-nology or format even comes close.

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. If there is data still extant in these formats 500 years from now, it would be a relatively trivial matter to access that data, given the onward march of technology and the world-wide adoption of these technologies at this time.

## 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.


## Author Biography

***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*