

Kathleen Maguire
Video Restoration
11/21/2007

Storage Options for Digital Files

In order for Deep Dish to adequately preserve the data and content created during the production process the materials must be transferred to a storage space and format that is conducive to long-term stability. Unlike other audio-visual formats, there are currently no consistent standards for long-term digital storage and preservation. There are several viable options currently in use in the field. Each option boasts attributes that may benefit the needs of Deep Dish, but also display certain drawbacks that must also be taken into consideration during the decision making process. Ultimately, the choice Deep Dish makes in regards to long-term storage and back-up will be based on a number of internal variables, with budget availability figuring heavily in the final decision.

Storage options:

Although standards are not established, it is well established that the key to successful preservation is through a system of back-ups created in multiple formats and stored in various physical locations. In the digital age the mantra is “L.O.C.K.S.S.”: lots of copies keeps stuff safe.¹ While this recommendation of multiple copies in a variety of formats is the best practice, it is also understood that such an extensive system is costly and thus not viable for many institutions. Outlined below are multiple options that could successfully maintain an organization’s files for the long-term.

Choosing the adequate storage environment and level of back-up is dependent, as stated above, on the available budget. However, the amount of materials Deep Dish plans to store, the type of materials, and the intended future use of the materials are also

¹ L.O.C.K.S.S. Homepage. <<http://www.lockss.org/lockss/Home>>. Accessed 11/7/2007.

significant factors for consideration during the decision making process. Deep Dish must seriously consider what their needs are within the organization in order to choose the most effective long-term storage option.

Linear Tape Open (LTO) –

Linear Tape Open is a magnetic tape format that works with an electromagnetic deck that writes to the tape as well as plays it back. With LTO, writing to the tape is conducted through a stationary head that runs the length of the tape, filling data bands one at a time. There are other digital tape options, such as advanced intelligence tape, digital data storage, and digital linear tape, however LTO has become the preferred tape option for archival purposes.²

The technology for LTO was created by Hewlett-Packard, IBM, and Seagate and is an, “open-standards, licensed technology.”³ The format is available for licensing to interested companies who can then manufacture and sell LTO cartridges and associated drives. In order to license the technology, companies must sign a license agreement and comply with extensive standards set forth by LTO. Many major companies, including Fujifilm, IBM, Maxwell, Otari, Sony, Tandberg, and TDK, have licensed the technology.⁴ Although it can be argued that there are proprietary elements to LTO, forcing licensees to comply with standards eradicates many of the problems generally associated with proprietary formats. Compliance promotes interoperability and develops

² Bigelow, Stephen. “Tape Backup Overview.” [Search Storage.com](http://searchstorage.techtarget.com/originalContent/0,289142,sid5_gci1165427,00.html). <http://searchstorage.techtarget.com/originalContent/0,289142,sid5_gci1165427,00.html> 7 August 2007. Accessed 11/10/2007.

³ Cholia, Shreyas and Nancy Meyer. “A Beta Test of Linear Tape-Open (LTO) Ultrium Data Storage Technology.” [National Energy Research Scientific Computing Center](http://www.nersc.gov/projects/probe/IBM_LTO_test_1.pdf). <http://www.nersc.gov/projects/probe/IBM_LTO_test_1.pdf> Dec 2001. Accessed 11/10/2007.

⁴ Linear Tape Open. “Licensing.” <http://www.lto.org/licensing/default.php?section=1&subsec=default>> Accessed 11/11/2007.

a much larger pool of interchangeable playback devices.⁵ Currently, there are four generations of LTO available and generations 5 and 6 are in planning stages. Generation 4 is capable of storing up to 800MB of 2:1 compressed data. Generation 6 is expected to be capable of handling 6.4TB of uncompressed data and to transfer at 540MB per second. In an archival environment, it is expected that most transfers will be uncompressed. Generation 6 is would have a capacity of 3.2 TB transferred at 270MB per second for uncompressed data.⁶

Currently, LTO's 2, 3, and 4 are commonly used for archival purposes. There are many different companies that design and sell systems to manage and automate LTO storage and back up. Requiring only a small amount of physical space, these systems, or tape libraries, can hold many LTO cartridges and also multiple LTO drives. An organizations' storage needs will determine the amount of tapes a library will be capable of holding. Systems range from the capacity to hold several tapes, to the capacity to hold hundreds. LTO tape libraries are then capable of reading from and write to the many cartridges stored in the library without physical supervision and based on predetermined automated settings and back-up intervals.⁷ Tape libraries often include many features that maintain easy access to the stored materials and exceptional control over the system. For example, Qualstar's TLS series features "Q Link," which enables a user to have remote access to the library and grants the ability to, "configure, upgrade, and monitor any TLS-Series tape library via a company intranet or over the internet." These models also include a "logical library" system that allows multiple computers to share a single tape

⁵ "What is LTO?" The Tech-FAQ. <<http://www.tech-faq.com/lto-tape.shtml>>. Accessed 11/6/2007.

⁶ LTO. "LTO Background." <<http://www.lto.org/About/ltoBg.php?section=3&subsec=ltoBg>>. Accessed 11/11/2007.

⁷ "TLS and RLS Series Tape Libraries." Qualstar. <<http://www.qualstar.com/146020.html>>. Accessed 11/13/2007.

library, both for back up and access.⁸ This level of management and operability is common among most manufacturers high quality tape library systems.

If an LTO based tape library has more storage than Deep Dish needs, an LTO based autoloader system may also be considered. An autoloader is comprised of a single LTO drive and holds a limited number of cartridges, usually 8 to 16, although some cap out at 20.⁹ An autoloader does not offer the extensive storage space and advanced operability available in LTO libraries. However, it is an automated system with significant storage space. The organization could also choose to purchase an individual LTO drive and perform all back up manually. The human handling of the back-up materials, however, increases the risk of error and damage.¹⁰

The costs of libraries, drives, and cartridges are dependent upon the LTO generation each product relates to. A generation 2 tape library that can hold 9.6TB of uncompressed data currently costs approximately \$8,000.¹¹ Autoloaders capable of handling up to 1.4 TB of uncompressed data retail around \$3,500.¹² Individual LTO-2 drives currently cost close to \$1,500.¹³ After this initial outpouring of funds, an LTO

⁸ “TLS-8000 Tape Library Systems with LTO Technology.” [Qualstar](http://www.qualstar.com/146081.html). <<http://www.qualstar.com/146081.html>>. Accessed 11/13/2007.

⁹ “TLS and RLS Series Tape Libraries.” [Qualstar](http://www.qualstar.com/146020.html). <<http://www.qualstar.com/146020.html>>. Accessed 11/13/2007.

¹⁰ Tandberg Data SMB Guide to Backup Best Practices. [Tandberg](http://www.exabyte.com/support/online/documentation/page_view.cfm?f=BackupGuide_1.html&p=PDF%20to%20HTML/basicbackup). <http://www.exabyte.com/support/online/documentation/page_view.cfm?f=BackupGuide_1.html&p=PDF%20to%20HTML/basicbackup>. Accessed 11/13/2007.

¹¹ Various consumer purchasing sources: <http://www.bizrate.com/tapecartridgedrives/products__keyword-lto2.html>, <<http://www.buy.com/retail/product.asp?sku=202476867&listingid=11646984&dcaid=17902>>. Accessed 11/13/2007.

¹² Various consumer sources. <<http://www.pcmall.com/pcmall/shop/detail.asp?dpno=648126&Redir=1&description=Tandberg%20Data-Magnum%201x7%20LTO-2%20Tape%20Autoloader-Tape%20Automation%20Libraries/Changers>>, <http://www.harddisk.com/tapedrive/library/powervault_122t_lto1.html>. Accessed 11/13/2007.

¹³ Backup Works.com. <<http://www.backupworks.com/index.asp?PageAction=VIEWCATS&Category=88>>. Accessed 11/13/2007.

system becomes relatively low in cost, which is one of its greatest appeals, particularly in comparison to hard disc drives and Digibeta's. An LTO 2 cartridge capable of holding 200GB of uncompressed data retails for approximately \$30. An LTO 4 that holds 800 GB of uncompressed data retails for approximately \$120.¹⁴

Hard Drives and Server Systems:

Hard disc drives are considered an important piece of effective digital back up. Currently, drives are costly. However, the cost is continually falling. Furthermore, HDD's storage capacity is continually increasing as their shell size is maintained, if not decreased. In early 2007, Hitachi released a 1TB HDD, which was followed in July of that year by 1TB HDD's from Seagate and Samsung. These 1TB HDD's began retailing around \$400 each.¹⁵ While LTO is cheaper option, if the organization is in the financial position to do so it is desirable for a complex system of HDD's to serve as the primary source of storage, with LTO acting a supplementary back up for access in the event that an entire system of drives fails. External hard-drives are well-established and ubiquitous devices that have permeated many markets. This minimizes the risk of obsolescence. Also, drive specifications are standardized and the major manufacturers conform to these standards, which helps avoid incompatibility and proprietary issues.¹⁶

A major concern with using HDD's for archival storage is that because most archival materials are only rarely accessed it is possible that continued lack of use of the drives could result in failure. To avoid this and properly maintain the drives, it is

¹⁴ Sortprice.com. <http://www.sortprice.com/search-CQ-Computer_Tapes-LTO_Cartridge>. Accessed 11/13/2007.

¹⁵ "Seagate, Samsung begin to ship 1 TB Desktop Hard Drives." *Daily Tech*. 19 June 2007. <<http://www.dailytech.com/Article.aspx?newsid=7740>>. Accessed 11/10/2007.

¹⁶ Wheeler, Jim. "Archiving on Hard Disk Drives." Joint Technical Symposium 2004. <<http://www.jts2004.org/english/proceedings/Wheeler.htm>>. Accessed 11/10/2007.

suggested they be run 2-3 times a year. To further safeguard the materials, files must be continually migrated to new drives

Despite this persisting concern, most users have found HDD's to be quite reliable and have only rarely experienced drive failure. In a presentation at the 2004 Joint Technical Symposium, Jim Wheeler noted that the hard drive failures he had experienced were only with HDD's manufactured by smaller manufacturers, suggesting that purchasing from one of the four major manufacturers reduces the risk of failure. Seagate, Samsung, Western Digital and Hitachi comprise the major four mentioned by Wheeler. Linda Tadic of ARTstor shared on the Association of Moving Image Archivists listserv that she found, of 200 hard disk drives used at ARTstor only four had failed. While this is a small failure percentage, had ARTstor not had in place an extensive back-up system that included multiple drives and alternate formats, the materials would have been lost.¹⁷

While it is understandable that even a rare failure may be a cause for concern at an organization, HDD's attributes and prominence in the field ultimately overcome this drawback. A major asset of HDD's is that they are, as noted by Jim Wheeler, "self-contained."¹⁸ An individual HDD can be immediately accessed by being attached to computers and systems already in place at an institution. Individual drives are also transportable which, while not conducive with archival practice, is beneficial in creating multiple copies of a file on multiple computers.

As with LTO's, it is possible - and preferable - to employ an automated Digital Asset Management System (DAMS). This system can either be outsourced to one of many companies who specialize in DAMS for HDD's or, if the demand for storage is

¹⁷ Tadic, Linda. < <http://palimpsest.stanford.edu/byform/mailling-lists/amia-1/2006/07/msg00006.html>>. Accessed 11/7/2007.

¹⁸ Wheeler, Jim. <http://lsv.uky.edu/scripts/wa.exe>. Accessed 11/7/2007.

great enough, run in-house.¹⁹ A drive based system operates similarly to a tape library with a system of drives onto which data is transferred to and stored automatically based on pre-arranged options and setting. The difference is that the data is stored on a HDD (there are also other options), as opposed to tape.

An option in server storage commonly applied in archival settings is use of a RAID system, redundant array of inexpensive drives. Like LTO's, there are a variety of levels of RAID configurations. With the exception of RAID-0, each of the levels commits data to multiple drives and boasts high data transfer rates. The advantages of a RAID system, particularly over individual HDD's are clear. In the system, an array of drives is presented as a single, accessible, drive to its host computer. This allows information that appears on all drives to be accessed from a single location without the instability of the single location acting as the only point of storage.

The system is arranged in a hierarchical manner with physical drives, physical arrays, logical arrays and logical drives. Physical drives are arranged into physical arrays, which are a group of drives that are either striping or mirroring information. A logical array will usually correspond with a physical array, and a logical drive will correspond with a single logical array. This complicated system describes how the data is communicated between the RAID system and the users computer. The physical array is stored in the system, while the logical array is the way the physical arrays are communicated to the computer to result in human readability. An early paper proposing the system notes that each group has, "Extra 'check' discs containing redundant information. When a disc fails, we assume that in a short time the failed disc will be

¹⁹ Tadic, Linda. "Toward a Digital Code of Hammurabi." Performing Arts Resources. Vol. 24, 2004.

replaced and the information will be reconstructed onto the new disc using the redundant information.”²⁰ This back-up information, which is sent to a randomly chosen drive, works to protect data on a higher level.²¹ RAID systems can be employed through either software or hardware, however software systems are often lower performance and require a higher level of user expertise.²²

RAID systems employ two types of redundancy measures, striping or mirroring. Striping divides files into several chunks and disperses individual chunks of data to different drives. For example, a 50KB file in a 5 drive system will have 10KB of data on each of the drives. Should a drive fail, there will still be data loss in this system, and thus the RAID level that rely entirely on striping, levels 0 and 5 are not the best solutions in an archival setting. Mirroring is a system in which complete files are backed onto two drives that read as a single drive in your system. This system is beneficial for archival purposes because two copies of data are consistently available. The drawback to mirroring, which is used in level 1 RAID systems, is that it requires twice the amount of storage space.²³ Level 10 RAID employs both options by mirroring all data and then striping it across all discs. Of the RAID levels this is the most reliable, however it is also the most expensive.

The costs of RAID systems vary greatly. It is possible to build one’s own system for a fairly moderate price, however, the cost of a decent and complete system can be expected to be similar in price to an LTO tape library. Where a RAID system becomes

²⁰ Patterson, David A., Garth Gibson and Randy H. Katz. “A Case for Redundant Arrays of Inexpensive Drives.” <<http://www.eecs.berkeley.edu/Pubs/TechRpts/1987/CSD-87-391.pdf>>. Accessed 11/14/2007.

²¹ “RAID Systems.” US Byte. <http://www.usbyte.com/common/raid_systems_2.htm>. Accessed 11/10/2007.

²² Pons, Juan A. “The Ultimate Backup: RAID for Your Home Computer.” <http://www.naturescapes.net/032004/jp0304.htm>. Accessed 11/14/2007.

²³ Pons, Juan A. “The Ultimate Backup: RAID for Your Home Computer.” <http://www.naturescapes.net/032004/jp0304.htm>. Accessed 11/14/2007.

much more expensive than LTO is in the drive cost. One TB of storage can cost over \$1000.²⁴

An organization may chose to implement an unautomated hard drive back-up system comprised of individual drives not working from a central location or system. Ultimately, however, this is likely to require much more work than a RAID system and is also much less stable. Such a practice is highly discouraged.

Output to Tapes:

Reformatting to Digital Betacam tapes has become a best practice standard for the preservation and safeguarding of video materials in archival and library settings. The format boasts high integrity and superior image quality. Digibeta is prominent throughout the professional broadcasting industry, which ensures extended availability for the format and playback devices. Sony introduced Digital Betacam in 1993 and has since continued to upgrade the product, with a 3rd generation of the associated camcorders released in 2005/2006.²⁵

Digitbeta has proven to be a stable format that, when transferred properly, contains little loss. The tape is specifically designed to avoid oxidation and maintain proper image quality for the long-term. Despite the rarely well-established quality of content maintenance available through Digital Betacam, it would be a major step backwards to begin implementing long-term storage with Digibeta as the primary preservation and back-up source. As pointed out by R. Justin Dávila, who worked as a system architect on the SAMMA system, “the tipping point where it is more effective to

²⁴ “RAID Storage: SATA Arrays.” RAD Direct. <<http://www.rad-direct.com/ProdFam-RAID-Storage.htm>>. Accessed 11/10/2007.

²⁵ “Digital Betacam Family 2005/2006.” Sony. <[http://assets.sonybiz.net/doc/bv/X/DigitalBetacam\(brch\).pdf](http://assets.sonybiz.net/doc/bv/X/DigitalBetacam(brch).pdf)>. Accessed 11/9/2007.

archive video as data rather than as video is well behind us.”²⁶He goes onto note that, often the difficulty for archives is not deciding whether or not to switch to digital storage, but deciding how. Since this project is initiating long-term storage at Deep Dish, it is suggested that high-cost Digibeta back-up tapes be avoided in lieu of a more cost friendly and forward-looking option of digital storage.

Two of the major drawbacks of Digibeta tape are the cost and workload associated with reformatting. A single 124 min Digital Betacam tape costs \$60, which is significantly more expensive than previously discussed data storage options.

Transferring to DigiBeta’s also requires either a proper in-house lab set-up or costly outsourcing. If in-house reformatting is established, there must be properly trained staff responsible for achieving the intended quality of transfers. The process of transferring to Digibeta is done in real time, with each transfer requiring some level of supervision for the entire length of the tape in addition to set-up time. Ultimately, this process is a much larger drain on financial and work-power resources than other automated options.

Furthermore, transferring to Digibeta is slowly becoming an outdated best practice that will likely be a rarely used preservation format in the near future.

Conclusion:

Since Deep Dish is beginning to implement a storage system it is suggested, if the budget is available, for a RAID system to be employed. If the organization is unable to afford a suitable RAID system, it is suggested Deep Dish look to LTO option as opposed to developing a shoddy or incomplete RAID option.

²⁶ R. Justin Dávila. “Archiving on video on HDD?.” [Amianet.org](http://sv.uky.edu/scripts/wa.exe?A2=ind0511&L=AMIA-L&P=R21912&I=-3). <
<http://sv.uky.edu/scripts/wa.exe?A2=ind0511&L=AMIA-L&P=R21912&I=-3>>. 21 Nov 2005.

