Backup Solutions

'Backup solutions' for computer data are preventative Information Technology (IT) strategies typically employed to mitigate both: i) the loss of, and damage to, data; and, ii) the amount of downtime for functionality, in the event of computer systems failure. (These two liabilities are often referred to, respectively, as the Recovery Point Objective (RPO) and the Recovery Time Objective (RTO).[i])

Approaches to backup solutions involve both software and hardware—the former in order to duplicate data from individual computers within a given networked computer system, and the latter in order to physically store the duplicate data somewhere where it will not be easily accessed or changed by a non-administrator. The architectural planning and administration of computer systems and networks directly dictates backup solutions' effectiveness and success rate, and will involve some inter-related structuring of both hardware and software.

This paper will provide an overview of several backup solution hardware and software, compare the strength of each in terms of the Library of Congress' Sustainability Factors, and provide specific comparisons with practical considerations for moving image collections and digital repositories. Prior to some of the specific software solutions discussed in this paper, basic backup utilities such as TAR and DUMP and NTBACKUP have been widely available and implemented, yet offer a low level of sophistication and are currently mainly used in noncommercial computing settings; they will not be discussed.

I: Concepts

It will be helpful to outline a few basic terminologies and core concepts, before comparing hardware and software.

Full, Incremental, and Differential Backup

W. Curtis Preston, a backup and IT professional who runs <www.backupcentral.com>, categorizes varying extents of data backup, into these 'levels':

> "FULL/LEVEL 0 Backup: full backup
> LEVEL 1: incremental backup of everything that has changed since last LEVEL 0
> LEVEL 2-9: each level backs up whatever has changed since the backup of the next lowest level; a level 2 backs up everything that has changed since a level 1, or since level 0, if there is no level 1.
> INCREMENTAL: a backup of everything that has changed since a backup of any type"[ii]

As 'levels,' the point in time each data duplication 'event' occurs directly dictates its authority as 'correct data.' The relationship of specific physical volumes of duplicated (aka 'backed up') data will continually vary, in terms of authority, with a computer system/network's most recent data state. A less authoritative level/backed up tape copy of data at one point in time may, at the same time, prove too old a snapshot of data yet also be crucial to the restoration of a system/network in the event of a failure down the line.

'Differential' backups refer to those, which chart all of the incremental changes made to a client, or system, since the last full backup.

For data security and integrity, a full backup is the preferred method of backup, however full backups require large amount of downtime (backups are usually performed at night, when client usage is limited or none) and storage resources. They are not usually performed more frequently than on a daily basis (if that often).

In the realm of backup, however, author Preston warns against reliance on too complex an incremental approach: "Backing up everything is easier and safer than backing up from a list," and, "I think that not backing up everything is dangerous."[iii] Preston further advocates full backup when he writes, "the more complicated your backup system is, the more likely it is to fail…Remember, special is bad."[iv] Preston's words are prescient when considering the promotion of data de-duplication features offered by commercial backup vendors. In one recent Symantec promotional video, data de-duplication is touted as a means of 'putting data on a diet,' with a Symantec spokesperson encouraging customers to stop purchasing storage space and instead employ Symantec's de-duplication software.[v]

Furthermore, this paper does not address 'snapshot' backup—merely a record of the state of a computer system's configuration. A snapshot may give an indication of how a system's data was configured at in use at any point in time, aiding in a diminished RTO and more recent RPO, however a snapshot does not backup data.

Failure

What is failure? System failure (aka an 'outage') can occur in a variety of ways, due to either equipment or human failure. Disks can fail, as can servers, computers, data tape carriers, and system software. Humans can also fail—making mistakes, in spite of functioning equipment performance, and introducing corruption or accidental data loss due to errors in system administration. Humans can fail to properly execute software commands, resulting in inadequate backups. Humans and equipment can also collaboratively fail, as might be the case in the demagnetization of data tape due to exposure to a magnetic field, the introduction of a virus into a system's backup architecture, or the dropping of a portable disk drive with backed up data on it.

Failure can be quickly remedied, or it can be disastrous; it can cause minor interruptions in network service, easily restored once a minor data restore is performed, or failure can mean the complete and irreversible loss of data in worst cases of negligent planning and maintenance.

Testing

In spite of the extreme undesirability of real-life situations where such attributes will be proven, full system testing is expensive and challenging to regularly test.[vi] As Klaus Schmidt notes, in his 2006 book *High Availability and Disaster Recovery*, Disaster Recovery, "is rarely utilized, and is usually tested infrequently."[vii] Simulating a partial or bare-metal restore requires system downtime and can be seen as risky and potentially data-endangering (while, ultimately, in the best interests of safety).[viii] As New York University Digital Library staff member, Brian Hoffman, states: "It's usually not going to be evident that you have a good or bad Disaster Recovery plan until something happens." Hoffman adds, "I think that, like the stock market, companies are short-changing their long-term interests in order to have a better next quarter. These people might question the value of sinking money into a .05% chance of something bad happening."[ix]

The cost-cutting aversion to full-scale testing of a system's backup functionality is reflected in W. Curtis Preston's comments on the tendency to rely on partial backup: "Backing up selected drives of filesystems is one of the most common mistakes that I find when evaluating a backup configuration. It is a very easy trap to fall into because of the time that it saves you up front."[x]

Yet, testing of the efficacy of backup software and hardware should be considered an integral part of managing any digital repository according to the OAIS model of ongoing management. Klaus Schmidt observes, "simply adding a backup component is not enough. This backup component needs to be *managed*."[xi]

Enterprise

One final note on common backup concepts regards the term, 'enterprise.' Usually, this refers to a large-scale system with multiple servers, multiple clients, multiple client-operating softwares, and diverse usage levels. NYU's Brian Hoffman regards this term, when used to describe capability of backup software, as code for, "we can do it; no job is too big."[xii]

II. Backup Solutions: Hardware

Tape

The roots of storing information on magnetically charged physical carriers reach far back into the nineteenth century, with Oberlin Smith first conceiving of and patenting the idea in 1878.[xiii] Though Smith's implementation, using magnetized wire, wasn't translated to plastics until World War II, the technology's eventual 'tape'-based commercial acceptance radically changed moving image and audio industries (not to mention amateur cultures). While the future creator-use of the technology in those predominant two realms of media appears decidedly in an irreversible death spiral, magnetic tape is still a widely used storage format for computer data.

Below are brief overviews of the two currently most prevalent data tape formats—both of which are the descendants of competing companies' previous technologies: IBM's ½-inch 3480 tape format, and the Digital Equipment Corporation's DLT format. Both histories and technologies can be considered intertwined.

Across tape storage formats, the self-contained cartridge prevailed over older reel-to-reel tape formats, and operates in tandem with corollary tape cartridge drives. Data is written in a serpentine line in byte-level data 'blocks' on separate tracks on the magnetic tape. This 'block' method does not allow for the alteration of data within each block; instead, changes to data within discrete blocks require the re-writing of the entire block of data. In the drive for higher storage capacities, the prevailing strategy has been for tape tracks to become smaller and smaller in size. Because of the size limitations and the reality of their physical nature, data tape formats have been developed hand-in-hand with a variety of robotic 'tape loaders' or 'autochangers' with the goal of increased automation.

DLT (and SDLT)

Described as, "Today's Hottest Storage Technology," in the 2001 *Quantum DLTape Handbook*, Digital Linear Tape (DLT) is a ½-inch magnetic tape cartridge format used for

data storage.[xiv] It was first developed by the Digital Equipment Corporation (DEC) in 1984, and later acquired by the Quantum company in 1994 who continued its development. Prior to the introduction of its LTO competitor, DLT storage was the dominant standard for small and middle-sized systems; after the introduction of LTO, Quantum introduced an updated SDLT version/format.[xv] Since its introduction, DLT technology advanced in its configuration of writing up to 94 MB on 22 tracks, to a potential of 800 GB on 448 tracks on SDLT formats.[xvi] Across DLT's numerous versions—including the SDLT line—its tape drives has primarily used the SCSI port interface (a reflection of the format's DEC heritage). The Quantum company has no further plans to develop future versions of DLT and, via the acquisition of original LTO consortium company Certance (originally a division of Seagate), has joined the consortium of LTO manufacturers effectively signaling the format victory of LTO over DLT. In its role in the LTO consortium, Quantum has made certain LTO versions to be compatible with older DLT formats to encourage DLT users to migrate their backed up data to LTO.

LTO

LTO is currently the most popular ½-inch magnetic tape cartridge format used for data storage.[xvii] The underlying principle in the development of the LTO format was that it would be an 'open format' technology developed by a consortium of developers, and readily licensable to interested manufacturers. The initial consortium members were IBM, Hewlitt-Packard and Seagate (later spun-off as, Certance, eventually acquired by Quantum) and the first version, LTO-1, was released in 2000.

Several core principles derive from the LTO consortium's 'open format' planning. First, consortium members sought to plan subsequent version releases of LTO (a "Six-Generation Roadmap"), and to make versions of the format backwards-compatible for two previous generations of the format. Thus, LTO-2 tape would function in an LTO-4 drive. Second, LTO would not be a proprietary format so far as there existed only one manufacturer. Instead, the technology would be licensable to interested manufacturers who could manufacture their own tape and drive brands of the technology. The logic in this decision spreads the technology across several companies, as a way to mitigate obsolescence and stimulate both a competitive and standardized market for data tape. The LTO consortium is generally considered as a success, and the format seems to have generally supplanted DLT as the leading data tape format.[xviii]

Some other features LTO has over the older DLT format are its ability for variable speed drives, enabling data to be written at different speeds which adds flexibility to backup schedules, and its WORM (Write Once Read Many) capability, which protects data from being overwritten by eliminating tape's 're-write-ability.' This WORM feature is a relatively new feature on data tape employed as a means of dummy-proofing full backups intended for longer-term backup purposes.

Optical Discs

While employed at smaller scales of data storage as a useful backup hardware, optical discs such as CD-R and DVD-R remain primarily a backup hardware solution for the consumer market. Their physical instability and susceptibility to irreparable, whole-scale damage are reasons for this. The advent of higher capacity optical disc technologies, such

as Blu-Ray were speculated to increase the popularity of optical discs as a storage hardware, but that remains to be seen as widespread adoption of that format has yet to occur. Additionally, the popularity of spinning hard disk drives and the absence of the need to physically handle backup hardware may suggest that Blu-Ray will be used as a backup hardware only in a limited, smaller-scale, capacity.

Spinning Disks

Spinning disks, or 'hard disk drives' (HDDs), are self-contained circular magnetic disks to which data is written. As with magnetic tape, hard disks have over a half-century of implemented history beginning with the IBM 350 storage unit in 1956.[xix] The internal architecture of HDDs resemble somewhat that of phonograph turntables, with a moveable protruding arm that reads information contained on the spinning disk, though the HDD's employs a magnetic head in place of a stylus.

One feature of HDDs that is often deceptively touted as an end-all solution is the ability to store data in a RAID (Random Array of Independent Disks) configuration—essentially disk mirroring that will, theoretically, provide a redundancy of data should one of the disks fail due to a physical malfunction. RAID arrays can be configured in a variety of ways (RAID levels, RAID-0 through RAID-6) with the more complex RAID-6 configurations offering a greater distribution of data across different disks. While RAID technology can offer a valuable duplicated distribution of data, in the case of a physical disk failure, RAID can also be deceptively comforting. The disk mirroring strategy employed by RAID can only protect against hardware errors so long as they do not occur on multiple disks, and RAID does not offer much protection against software or operating system errors which may be merely replicated on RAIDed disks, in the course of normal mirroring. It is for these reasons that RAID protection is seriously different than backup, even though some may use both concepts interchangeably.

Hardware Comparison Across Sustainability Factors

Just how do the above hardwares compete in terms of the Library of Congress' (LoC) Sustainability Factors of: disclosure, adoption, transparency, self-documentation, external dependencies, impact of patents and technical protection mechanisms?[xx]

In terms of disclosure and self-documentation, LTO may hold an edge over its competitor DLT (and other revamped data tape formats such as Sony's SAIT) by virtue of its spreading of technology across several manufacturers. Though, it is unlikely that any recipes for tape binder will be offered to the public—a weakness in transparency shared by all of the above hardware options. With the issue of proprietary technologies, such as in the case of data tape and optical discs, sustainability will necessarily be compromised. Indeed, the external dependencies for hardware solutions are significant as commercial manufacturers hold the key decision making power to any formats continued production. Spinning disk hard drives may be the most implementation-neutral of the four, capable of use in a variety of computing situations, suggesting themselves as the most sustainable of the hardware options. However, despite the advantages their closed-system nature—namely, having the least opportunity for physical human interaction with the functional parts of the hardware (see: scratches on optical discs; dust introduced onto tape)—HDDs make minor repairs and adjustments less possible for the average user.

In addition to the LoC's sustainability factors, W. Curtis Preston advocates considering other factors in evaluating hardware solutions, such as: reliability, flexibility, and removability. Removable HDDs may offer flexibility in configuration and removability, but their reliability is hardly a sure thing even in RAID array configurations. Data tape and optical discs may prove somewhat more reliable if left untouched on a shelf for several years, but both are often singularly dedicated to a specific client or use.

Hey What about Cloud Storage?

'Cloud storage' refers to an internet-based storage scheme, often involving a vendor who hosts and provides access to data (in this case, backup data) for a fee. The main feature of this is quick access and the alleviation of the need for physically storing backup hardware. Many vendors offer this service, including some who make backup software (like the EMC corporation), and others like the Amazon S3 service. Essentially, though, on the vendor's end storage considerations will involve above hardware options (unless the vendor somehow contracts out to another storage provider!).

III. Backup Solutions: Software

This section will consider two of the most popular Open Source backup softwares, Amanda and Bacula, as well as Symantec's proprietary Backup Exec software for small to mid-size system backups and Retrospect—the EMC corporation's proprietary backup software for larger, 'enterprise' systems.

Amanda

Amanda is the acronym for the Advanced Maryland Automated Network Disk Archiver—an Open Source backup software initially developed by the University of Maryland in 1991, and currently maintained by Jean-Louis Martineau at the University of Montreal.[xxi] Until the advent of fellow Open Source backup software, Bacula, systems administrator James da Silva claimed Amanda to be, "the most well-known Open Source backup software."[xxii]

One of the principal reasons for Amanda's excellence as a backup software is its ability to support the backup of complex client/server system architectures. As networked computing advanced in the 1990s, the industry witnessed the growth of non-mainframe computing. Computing systems grew de-centralized and "heterogeneous," with several different clients (ie. connected user computers). Networks in, say, one company or organization began to involve clients with multiple operating system platforms (eg. Microsoft Windows for administration computers; Mac OSX for graphics and creative computers; and, Linux for IT and database departments). The need to backup data across different networks grew more important and complex than ever.

Amanda is able to handle these differently configured clients and runs via a daemon—a program installed on each client computer in a system that enables the remote running of 'background' software by a system administrator. Amanda uses a daemon called 'xientd.'[xxiii] Via the daemon, Amanda's configuration software is able to automate the copying of each client computer's data to either data tape or "virtual tapes" (ie. spinning disk storage).[xxiv] A feature of Amanda is that simultaneous backup to both physical and

virtual tape is possible, creating potentially crucial data redundancy across backup hardware.[xxv]

Another major feature is that Amanda has no dependency on any specific proprietary device driver—it can accommodate any.[xxvi] This means that systems administrators need not worry about an interruption in supporting a client due to proprietary and versioning limits on the operating system of that client. Again, the advantages of the non-proprietary, open software model are on display.

Amanda backs up data as configured by its 'scheduler.' In configuring the scheduler, an administrator is able to stagger incremental backups of clients on certain backup tapes or disks, while simultaneously running full backups of other clients. Various rotation 'schemes' for backups exist, including incremental backup rotation (where oldest data is written over with the newest), and the Towers of Hanoi scheme (which involves a complex algorithm rotation where tapes correspond to specific system disks). In such a flexible configuration, administrators can develop a backup/dump cycle timeframe that will accommodate the system's normal rate of change for data. If data changes more frequently on certain clients than others, automated backups can also be configured to update faster-changing client data more often than that of slower-changing clients. In fact, any possible rotation schedule can be developed and contributed to the software. All the while, Amanda enables less periodic full backups to occur simultaneously if needed (such as a monthly off-site tape dump). Amanda stores catalog info about its backups (when performed, at which level, etc.) in a self-contained database—one of the weaknesses when compared to its Bacula counterpart, which supports separate SQL-based databases.

The Open Source provenance of Amanda is, obviously, another of its major features. It is free, and is financially supported through the free labour of its user community and an academic institution (see: Jean-Louis Martineau, above). This institutional support suggests a long-term sustainability for the software that other, non-affiliated Open Source softwares may not enjoy.

BACULA

Bacula is another Open Source backup solution software, initially developed by Swiss programmer Kern Sibbald, beginning in 2000, and released to the public in 2002.[xxvii] Bacula is currently in its third version, with Bacula's 3.0.3 release on October 18, 2009.[xxviii] In its very short existence, according to Open Source web directory Source Forge, Bacula has become the most popular Open Source backup software. [CITATION]

Bacula and Amanda share many features in common as backup software, including: backup hardware versatility (to tape, disk, optical disc, or any combo thereof); complete adaptability in the configuration of backup data onto said backup hardware (backup spanning multiple volumes, autochangers); graphical user interfaces (GUIs); and the ability to handle complete backup for multi-platform clients.[xxix] Like Amanda, Bacula runs via daemons installed on each client it is configured to backup.

Unlike Amanda, Bacula supports its catalog information (generated data about the scale, time and success of backups) in a separate database and currently supports three different SQL-based databases: MySQL, PostgreSQL and SQLite.[xxx] This fact means the existence of an external dependency, which Bacula's design approaches by making backup of the database an option.

According to W. Curtis Preston, "the advanced feature of greatest importance to most users is the ability to do bare-metal recovery of machines using Bacula."[xxxi] A bare-metal restore is necessary after a complete system failure where both data and operating software has been lost on all clients. Bacula enables users to not only mirror system and disk images (providing a snapshot of configurations in their pre-failure state), but also to re-establish software without original manufacturer boot discs and to recover lost data via the most recent full and incremental backups. Subsequent versions of Amanda have been written to accommodate bare-metal recoveries, as well.

Bacula is also one of the most widely used backup softwares because of its scalability—namely, its capability of backing up both smaller systems with few clients and large, 'enterprise' grade systems with hundreds of clients and more complex architectures.

Symantec's Backup Exec for Windows Servers ver. 12.5

Backup Exec for Windows Servers version 12.5 is the most popular Symantec server/network proprietary backup software for data. Users purchase a 'core license' for each server on which the software will be used, along with annual service packages with different tiers of support offered (eg. Monday through Friday support, or, round the clock support). A single license of the software and one year of support costs $1,162.62; purchasing a Backup Exec core license alone costs $945.00, but additional charges apply in order to reinstate technical support; annual full-support subscriptions cost $217.41 and include any software updates.[xxxii]

Backup Exec backs up and can restore data, but not software or applications. Backup Exec has the capability of writing data to either HDDs or data tape (LTO or DLT) external hardware on a schedule determined by users. Backup Exec can be implemented to employ an 'incremental' backup strategy, which typically backs up data at the end of each day. Backup Exec can also be implemented to run a more frequent 'continuous protection' backup strategy—backing up data constantly so that, in the event of an outage or system crash, Backup Exec can restore data as close to the moment (ie. the 'backup point') before system failure as possible. This feature is possible with Bacula and Amanda, but is touted as a RPO-maximizing feature of Backup Exec. Symantec representatives recommend that the incremental strategy and the continuous strategy can be installed to run simultaneously on the same server, as they merely require separate setup and installation.

Backup Exec needs to be networked with the Storage Area Network (SAN) to receive and retrieve data. Backup Exec offers a high level of granular restoration—a single file can be all that need be restored.

Symantec's Backup Exec System Recovery ver. 2010

Symantec's Backup Exec software is divided between a backup solution software for data and one for operating systems and software. Backup Exec System Recovery version 2010 is the Symantec software more oriented towards disaster recovery as it can recover an entire operating system (applications, software) and associated data. However, Backup Exec System Recovery is only capable of incremental backup, and doesn't allow for a granular degree of restore (ie. a single file). Backup Exec System Restore employs HDD storage and does not offer compatibility for data tape—a Symantec sales representative claims this is a forward-looking feature of the software, given the decreasing costs of HDD

storage and the company's perceived trend away from tape storage. It is usually used in case of the need for performing a bare-metal restore. It is recommended that the two Symantec products be used in tandem for optimum security—a recommendation that most of Symantec's customers heed.[xxxiii]

Both Symantec softwares offer GUIs for administrating backup, and email notifications of log information whenever a full or incremental backup is performed. (Amanda and Bacula offer GUIs, as well.)

When asked about what kind of 'guarantee' is offered with Symantec's software solutions, Symantec sales representative 'Matt' was taken aback. Despite the nature of the paid service his vendor/employer offers to customers, Matt waffled and stated that nothing was guaranteed in the realm of backup. Furthermore, while having extolled the benefits of HDDs, Matt stated that he was not permitted to make any further recommendations about backup hardware.

EMC's Retrospect

Retrospect is an enterprise-capable backup software available from the EMC corporation. It is sold in a variety of 'packages,' and prices vary according to system configurations; multi-server, multi-client licenses cost $1,669.00 with introductory licenses for a single client (ie. consumer-level) starting at $119.00.[xxxiv]

As with the other backup softwares discussed, the EMC corporation's Retrospect software uses a daemon background application, which needs to be installed on each client computer in a system (each client requires the purchase of a license). It share many other features in common with Bacula and Amanda, too, including: the ability of multiple simultaneous backups; a GUI with email backup log report notifications; a backup scheduler for full and incremental backups; WORM-capable writing to data tape; accommodating functionality for writing duplicated data to a variety of backup hardwares (data tape, HDDs, optical discs); and, the ability to perform bare-metal restores.

One feature of Retrospect is that it includes a byte-to-byte MD5 checksum for verifying media files' integrity. While this is not explicitly noted as feature of the Open Source software options, an Open Source checksum software can easily be built into the regular operating architecture of a computer system. Retrospect also offers encryption capabilities for backed up data (a recent added feature of some LTO tapes, as well).

Software Comparison Across Sustainability Factors

Just how do the above softwares compete in terms of the Library of Congress' (LoC) Sustainability Factors of: disclosure, adoption, transparency, self-documentation, external dependencies, impact of patents and technical protection mechanisms?[xxxv]

Open Source options such as Bacula and Amanda fulfill LoC's disclosure and transparency factors, in that they are fully self-documented via an array of online manuals, user forums and wiki-documentation resources. Both have relatively developed user communities and adoption, with Amanda having an institutional academic sponsor, and are widely implemented. As such, however, both are reliant on community members to add to the software over time—an external dependency.

Proprietary options are naturally weak in terms of disclosure, and long-term transparency. While companies have begun to provide documentation on their products (in

an attempt to compete with the openness of Open Source options' full documentation), this documentation will fall short of enabling users to add programmable changes. Proprietary options may be more widely adopted in a similar implementation, possibly suggesting a longer existence in to the future. They are also tragically weak in that their utility may decrease significantly once customer/users stop paying annual subscription fees for support and updates. This weakness alone calls into question the sustainability of Retrospect and the Symantec options. The encryption feature of Retrospect adds an additional layer of complexity and danger into long-term considerations. What it the backed up data persists but the encryption key is lost, or no longer supported by a bankrupt EMC in the future? While it may provide a degree of security against data-stealing, encryption may be in violation of W. Curtis Preston's mantra to keep backup systems simple.

III: Concluding Analysis

In closing, I want to consider some of the most important factors in selecting and managing a digital repository for moving image and audio collections, while mitigating risk as best as possible.

With any moving image collection of video files it is likely that while files may undergo iterative changes at the mezzanine or access level (such as adding slates and watermarks for implementation or licensing purposes, or creating new composite files of several clips in a new sequence), master preservation files should remain secured and untouched. Thus, while backup software schedules should still be frequent it is the full backup of data that will prove most crucial in any attempt to restore data.

In terms of hardware, LTO data tape seems like the preferred option in spite of startup and ongoing costs for the reason that the LTO format closely resembles physical moving image carriers such as videotape, and a moving image collections manager may naturally feel more comfortable administering a physical object. Additionally, resources for proper tape storage may already be at such a manager's disposal. While the trend towards HDD storage may prove more appealing, there may be an argument for choosing LTO for the basic paranoid reason of keeping magnetic tape technology somewhat alive. Ideally, some combination of onsite HDD and offsite data tape storage would be implemented with the former being used for shorter-term, incremental backups, and the latter as a fail-safe full backup, backup solution. Concerns over cost, and a contingency of whether or not an organization already has a reliable storage server may practically trump the recommendation for data tape. Additionally, in the tragically common instance of passive stewardship (due to negligence, ignorance, a lack of funding, or otherwise), it seems more likely that data may be recovered from antiquated LTO tapes than from hard drives, if both were left to sit on a shelf for a decade.

Indeed, the verification of authenticity of data over time will perhaps be of increased importance compared to the need to create a full backup on a daily basis. The need for a checksum, thus, is of paramount importance and collections managers should consider a backup solution that can incorporate a checksum software. Retrospect explicitly includes this, though Open Source software can be configured to incorporate these, as well. Alternatively, a checksum may be incorporated into a backup workflow before and/or after data has been written to hardware.

Though the importance of backup should not be discounted, the complexity of changes to said mezzanine and access files may be considered minimal when compared to an organization that creates entirely new data from scratch. Thus, a backup solution software touting an incredible RPO-capability may be a less important factor for a moving image collection, when compared with an enterprise-level multinational bank's need for one.

Specifically addressing software solution options, Open Source options Amanda or Bacula offer a clear edge over proprietary solutions. As funding for moving image collections (so often, non-profit enterprises) is often project-based, justifying subscription fees for proprietary backup solutions may not be realistically sustainable. Maintaining operational funding can be a challenge even in physical domains, making the argument for the need for ongoing fee-based vendor support. Furthermore, the size of data involved at most moving image collections may not warrant 'first in line' priority when it comes to support or customized adaptability from vendors dealing with the computer systems of multinational corporations. Though Open Source options may involve a skill-set and knowledge-base potentially most commonly available only to IT professionals, it is plausible that collections managers might develop a knowledge base to manage a Bacula or Amanda software once backup protocol has been set in place by a programmer or IT professional. Alternatively, an organization's IT department may be able to take on the backup responsibilities in tandem with those already in place for other data; however, it will be important for collections managers to be involved to some minimal extent in the ongoing maintenance of their collection's file content.

In determining a preference for either Amanda or Bacula software, a moving image collections manager should consider both which is more widely adopted (Bacula), and whether or not similar organizations have already begun employing the software (where specific needs for programming may already have been contributed to the software). Ultimately, given data's non-physical nature, considerations for hardware redundancy will likely outweigh short-term functionality concerns, making tape a potentially more attractive option than HDD storage.

---

[i] Klaus Schmidt, *High Availability and Disaster Recovery* (Berlin: Springer, 2006). Pg. 27.
[ii] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 28.
[iii] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 27.
[iv] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 41-42.
[v] *Put Data Storage on a Data Diet with Symantec Data Deduplication* Retrieved November 12, 2009 on < https://www4.symantec.com/Vrt/offer?a_id=82085>
[vi] Symantec, "Five Critical Recovery Flaws Your Last Disaster Recovery Test Missed" *White Paper: Disaster Recovery* April 2009. Retreived November 12, 2009 at < http://whitepapers.silicon.com/0,39024759,60389993p,00.htm>
[vii] Klaus Schmidt, *High Availability and Disaster Recovery* (Berlin: Springer, 2006) Pg. 28.
[viii] Klaus Schmidt, *High Availability and Disaster Recovery* (Berlin: Springer, 2006) Pg. 28.
[ix] Interview with Brian Hoffman. November 16, 2009.
[x] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 27.
[xi] Klaus Schmidt, *High Availability and Disaster Recovery* (Berlin: Springer, 2006) Pg. 62.
[xii] Interview with Brian Hoffman. November 16, 2009.

[xiii] Friedrich Karl Enger (ed.), *Oberlin Smith as the Invention of Magnetic Sound Recording: An Appreciation on the 150th Anniversary of His Birth*. January 1990. Unpublished.
[xiv] Peter McGowan (ed.), *Quantum DLTape Handbook*. (Quantum publication: Eighth Edition, 2001).
[xv] Bob Doran, *History of Magnetic Data Storage* Retrieved November 25, 2009 on <http://www.cs.auckland.ac.nz/~bob/Old%20Books%20for%20Web/Computerhistory.html>
[xvi] Bob Doran, *History of Magnetic Data Storage* Retrieved November 25, 2009 on <http://www.cs.auckland.ac.nz/~bob/Old%20Books%20for%20Web/Computerhistory.html>
[xvii] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 272.
[xviii] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 272.
[xix] http://www-03.ibm.com/ibm/history/exhibits/storage/storage_350.html
[xx] *Sustainability of Digital Formats Planning for Library of Congress* Retrieved November 14, 2009 on <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>
[xxi] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 125.
[xxii] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 125.
[xxiii] *The 15-Minute Backup Solution: Secure Network Backups in a Heterogeneous Environment in the Time it Takes to Order a Pizza* Retrieved November 24, 2009 on <http://amanda.zmanda.com/quick-backup-setup.html>
[xxiv] *The 15-Minute Backup Solution: Secure Network Backups in a Heterogeneous Environment in the Time it Takes to Order a Pizza* Retrieved November 24, 2009 on <http://amanda.zmanda.com/quick-backup-setup.html>
[xxv] Amanda wiki. Rettrieved November 25, 2009 on <http://www.backupcentral.com/components/com_mambowiki/index.php/AMANDA>
[xxvi] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 127.
[xxvii] *Interview: Kern Sibbald* February 6, 2007. Retrieved November 25, 2009 on <http://archive.fosdem.org/2007/interview/kern+sibbald>
[xxviii] *Bacula Version 3.0.3 has been released to Source Forge* October 18, 2009. Retrieved November 24, 2009 on <http://www.bacula.org/en/?page=news>
[xxix] *Bacula vs Other Backup Solutions* Retrieved November 18, 2009 on <http://wiki.bacula.org/doku.php?id=comparisons>
[xxx] *What is Bacula?* Retrieved November 28, 2009 on < http://www.bacula.org/en/dev-manual/What_is_Bacula.html>
[xxxi] W. Curtis Preston, *Backup & Recovery* (Cambridge: O'Reilly, 2007) Pg. 172.
[xxxii] Telephone interview with Symantec sales representative. November 14, 2009.
[xxxiii] Telephone interview with Symantec sales representative. November 14, 2009.
[xxxiv] Product website. Retrieved November 18 2009 on <http://go.iomega.com/en-us/products/backup-software/retrospect-windows/download/?partner=4760
[xxxv] *Sustainability of Digital Formats Planning for Library of Congress* Retrieved November 14, 2009 on <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>