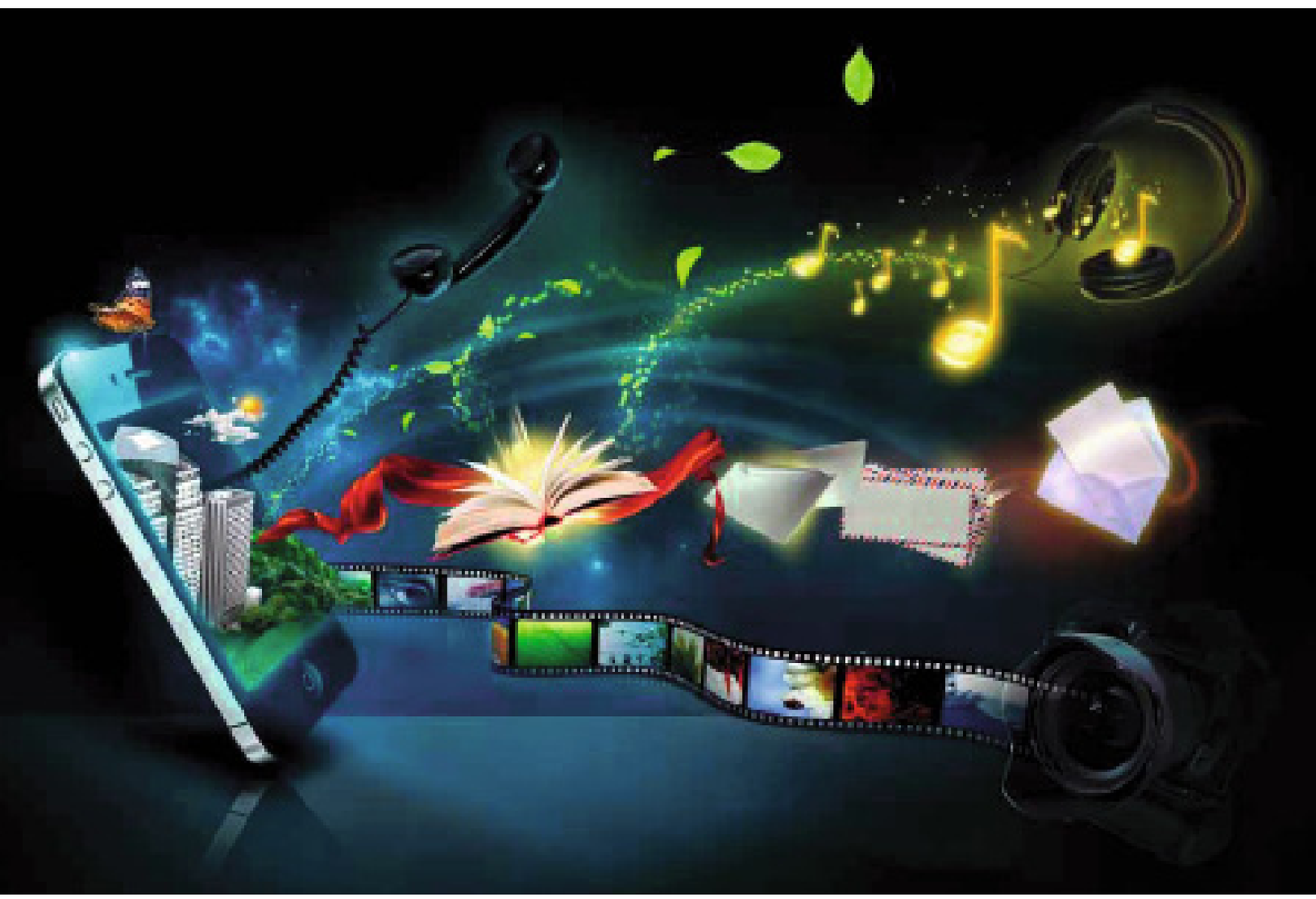
Data Sets Explode and Often have to be Secured without IT Department

Media Industry in a Squeeze

Many companies have been fighting the battle against an exponential increase in data volume for some time. The same goes for the topic of data security. The challenges that come with a growing amount of data are well documented in examples from the media industry: production formats are growing, so that video, photo and audio files are larger than ever before. In addition, the number of formats which are produced and supplied are growing as well. As the production cycle gets shorter, this results in an increased risk of user error and loss of data. With an adequate security strategy, it is possible to mitigate these risks.

A simple yet effective question is proposed at the beginning of every backup planning: “How long can we last without the data?”. The answer varies between different data sets. While time-critical data has a tolerance of only minutes, basic data could be offline for hours. There is some data (perhaps closed financial information) that could be inaccessible for days without impacting business negatively.



From this point of view it's possible to develop a classification that helps in the planning of data security. (Image 1)

Outside of the very time-sensitive data, most other production data can be secured with a traditional backup, respective to the restore process tolerance. Here, the restore time depends on the size of the file, taking between minutes and hours. Completed projects and business proceedings can be moved to an archive for storage. If the archived media is stored offsite, then it first has to be brought back to the office. This way, three clear segments are found which can meet your requirements. In the following paragraphs we'll shine some light on the individual solution possibilities.

Time-Sensitive Data requires data availability

The central storage in media production, usually in the form of a SAN (storage area network), is the starting point for all steps of production in order to offer maximum performance for multiple users simultaneously. Often times there are many coworkers working on different parts of a project at the same time. If this storage goes offline then production screeches to a halt. Costs continue to pile up, and due dates don't take delays in to account. A company's reputation is also on the line. As most of these circumstances are dealing with many terabytes of data, a restore can be expected to take a considerable amount of time, even days in extreme cases. This is where the limits of a traditional backup are reached.

A solution can be found in using cloning or mirroring technology. This enables a SAN or primary storage to be copied identically to a second storage (including Access Control Lists, or ACLs, extended attributes, extended Finder Attributes, etc.). This takes place automatically at regular intervals, or based on a structured schedule. In case of a problem, the administrator can make the second storage available, and work can be resumed within minutes. A restore process isn't needed, which is the epitome of a data availability solution. A popular solution in the production field, P4 Synchronize by Archiware, goes even a step further and offers data versions and snapshots for the cloned storage. Additionally File System Events are supported and can be used to secure large data sets more efficiently and achieve shorter sync intervals. (1)

Not all data is time-critical, which is why there is also room for a traditional backup to disk and/or tape. While storing on disk allows quicker access, storing on tape is much easier to grow with, cheaper, and easier to maintain with large quantities. (2) When using P4 Backup and P4 Archive the necessary tape library can be used for backup as well as for archive. Individual users can be set up with rights to restore from backup or archive, so that the administrator doesn't have to maintain this function, and data that was perhaps accidentally deleted can be restored directly.

Be without data for...

Minutes	**Data availability/Failover** NO restore necessary
Hours	**Backup** to Disk or Tape cyclic with Restore (and off-site storage)
Day(s)	**Long-term Archiving** with off-site storage

Image 1

Basic rules of data security

Tried and true rules should be adhered to, in order to accomplish the highest possible level of security. (Image 2)

Automatic Securing of Data

One of the most important steps in professional security is the automation of the process. Completing the process manually risks that the securing of data remains incomplete because something else appears more important, such as completing a production, or perhaps it is a holiday and there is no one to run the process. Only an automatic securing of the data can save the day in case of data loss.

Multiple levels of security

The increasing complexity of today's IT systems is often over-looked.
The more components in a system, the higher the chance of a problem. With multiple computers, servers, routers, operating systems, and processes there are unmanageable interactions buried deep under the surface. Data security should reflect this with multi-level security. The data availability solution should be rounded out with a Backup, the Backup with off-site storage, in which tapes can be employed.

Completeness

The question of what should be secured is key. Naturally, the server and its drives should be included. At this point it is important to reflect on the relevant configuration settings, such as volume- and IP-lists, router settings, and so on. Specialized workstations may be worth securing individually, given the highly individualized tools, plug-ins, and expansions. If external employees or freelancers are a big part of production, their mobile or external workplaces should be included in the plan as well. Since the backup of mobile workstations is a challenge, specialized solutions like P4 Backup2Go are to be preferred. The installers for all software implemented in production can be collected in a central location. Equally recommendable is documenting and securing the exact workflow of production and its parts. It's helpful to archive this documentation with each respective project in order to avoid confusion at a later time.

Changing the media

To limit technology-related risks, different mediums for saving data have been used. Not too long ago, optical media such as CDs, DVDs and others were a close second to hard drives. Today, however, the reliability, read/write rate, and capacity of these mediums just aren't sufficient. In fact, one medium is geared towards professional data security while holding larger amounts of data (along with disk), being LTO Tape (Linear Tape Open). The technology is being implemented around the world, from banks to insurers to Google, and smaller companies with smaller wallets can also implement it. The LTO-norm is a bit of a stroke of luck in the history of IT. The best features of previous technologies have been collected into a future-capable solution. LTO tape thus captures the best parts of the predecessors, without having the negatives. The LTO consortium is comprised of IBM, HP, and Quantum. Only these three manufacturers build drives that are then in turn used by other manufacturers for libraries. Performance and reliability of today's generation of tape solutions are impressive. Especially noteworthy is the storage life of 30 years. That means LTO tape is the only storage medium that can backup its claims with proven storage duration. (Image 3)

Relocation

Damage occurring locally can impact the whole IT structure. Often times, the securing of data that takes place in the same room or rack experiences the same fate. As data makes up the core of firms these days and are the most important factor to their business, an additional securing of data is necessary – a secure offsite copy. This step protects against major damage and catastrophes, and protects significant company data. A professional offsite storage solution is comprised of three sets of data: one set can be found at the company location, the next is kept offsite, and the third copy can be found in transport either to or from the company to before or after being updated. This ensures that damage incurred while in transport can´t destroy the complete backup. Where the offsite storage takes place is relatively unimportant, as long as it's a decent distance away from the company location, and it's a dry, dark and cool space. Theoretically, this could be a closet at the CEO's home. The more often the offsite copy is rotated, the lower chance of major loss of data. A unique form of offsite storage is keeping the copy or mirroring at a different location.

Basic Backup Rules

- Automatic Securing
- Multiple Levels of Security
- Completeness
- Changing the Media
- Relocation/Off-site Storage
- Restore Test

Image 2

LTO Advantages

	LTO 5	LTO 6
Speed	up to **140 MB/s** native	up to **160 MB/s** native
Capacity	**1,5TB** native	**2,5TB** native
Reliability	servo tracks, auto speed, verify, error correction	
Archive Life	certified for **30 years** (!)	
WORM	compliant + long-term	
Off-site Storage	simple & cost efficient	

Image 3

P4 Synchronize can be used to clone the data to the secondary storage that is far enough away. This option is somewhat expensive and complex, and implementation is often limited to certain scenarios. Cloud solutions are limited as a realistic alternative, given the numerous unpredictable variables. Upload rate, accessibility, and workload in the data center, as well as the performance when restoring are outside the personal influence, and are thus a game of luck. Larger amounts of data still require disproportionately long restore times. Equally unattractive still is the cost of such cloud based solutions.

Restore Test

A securing of data is considered complete only once a restore has been run to check the process. This point is often overlooked. Besides reading the data the restore process can be an organizational challenge. When a complete server or storage location goes offline, where can data be restored to? Where is the reserve capacity, backup computers, and necessary information for the restore? Only by testing the full procedure are all relevant steps understood. This test run should be diligently documented in order to provide step by step direction if needed. A good time to try this dry run is when integrating a new computer that can be used as a tester before being setup for its eventual role. Changes in the overall setup, network topography, and the utilized (backup-) software versions provide probable adaptations for the restore set up, and thusly the documentation. The optimal time for this is the purchasing of a new computer which will always be included in the data security steps, impacting in detail the securing of data, respective run times and capacity of the backup.

MAM as central piece

An increasing number of media companies use media asset management (MAM) solutions. This allows the management, cataloging, distribution or review and approval of all media. Tied in to the central storage, usually a SAN, the overview provided to all coworkers is the most effective solution for the hundreds of thousands, or even millions, of assets. The availability of the central storage is assured as described above. Additionally, integration of an archive makes sense in order to keep assets no longer needed off the expensive storage. For two of the most attractive MAMs, CatDV and CANTEMO Portal, integrations with P4 Archive are available. Users can thus archive from within the interface of the MAM. (3)

A professional level of security can be built by combining disk and tape elements. This creates the optimal performance and accessibility (on the disk side) plus long and affordable retention time (on the tape side). Sync-to-disc- plus Backup-to-tape (or Disk2Disk2Tape) combines the best of both worlds and implements the strengths of each medium. This way the primary storage can be cloned to a secondary storage, which then serves as the source for a tape-based backup. The tape storage allows extremely cheap storage combined with long storage durability, as tapes can be added to the library if needed. (4)

RAID is not Backup

It is often mistaken that using a RAID system can pass for data security. In truth, using a RAID increases the complexity considerably, as well as the likelihood of a problem. A RAID system protects against failure of one or a few hard drives, depending on the RAID level. There is no protection against user error, unintentional deleting of data, malware, viruses, or other risks. Limited protection through RAIDs is only accomplished through regular monitoring and a cyclical maintenance. In case of an emergency such as drive failure, the whole data set is at risk of being lost, because all data sectors have to be correct and available during a rebuild. A rebuild of a larger RAID can take many hours, up to multiple days. While this rebuild is taking place, access to data is limited and risky. Each RAID requires appropriate security. (5)

Backup and Archive in Comparison

The difference between Backup and Archive isn't easy to discern upon first glance. A Backup is always a copy of data that is used in production. These copies are regularly and automatically saved at a predetermined time and kept until the retention time is met. After that the backup updates itself overwriting previous copies. It is a cyclic process.

An Archive, on the other hand, is a long-term retention of data that is no longer involved in the production cycle. Often times a production will be archived manually once it has been closed or shipped. No longer needed on the central storage, the data is deleted from online storage. This process is called a migration, or moving from online to offline storage. Data kept in the Archive can't be easily deleted, and the Archive grows continuously with the collection of newly archived data. Because of the growing amount of data, it is important to have a plan to catalogue data, such as through Meta Data that describe the contents of assets, in order to find data quicker when needed in the future. P4 Archive by Archiware has the additional advantage of a browser interface that adds the flexibility to access the Archive from anywhere. (6)

A well planned securing of data nowadays is easily structured without an IT specialist. Easy to use software together with combinable functions geared towards accessibility, Backup and Archive are noteworthy criteria. Also necessary is a measured implementation of resources, and foresight in planning. The realization of such a solution will help all of those involved sleep a bit more assured, and braces the company against unwanted surprises and threats to its existence which can be brought on by data loss.

Dr.med. Marc M. Batschkus, Medical informatics specialist, at Archiware responsible for business development.

(1) http://www.andre-aulich.de/en/perm/use-xsan-2-2-xs-filesystem-events-and-archiware-presstore-4-x-to-protect-large-volumes
(2) White paper http://www.spectralogic.com/index.cfm?fuseaction=members.docContactInfoForm&DocID=4255
(3) http://provideotech.org/castor/
http://moosystems.com/products/moofs-archive-app/
(4) http://www.archiware.com/disk-to-disk-to-tape-d2d2t.33.1.html
(5) Paper in German that has all the formulas and as a result towards the end Mean Time To Data Loss with Bit Error Rate:
http://www.heinlein-support.de/upload/slac08/Heinlein-RAID_Mathematik_fuer_Admins.pdf
(6) http://www.archiware.com/video-archive.35.1.html