# How Do We Archive Our Digital Data?

By

Ray Quattromini – Managing Director

May 2010

Fortuna Power Systems Ltd

Fortuna Power Systems Ltd
1 Woodmere Croft
Basingstoke
Hampshire
RG22 5HB

T: +44(0)1256 782030
F: +44(0)1256 359969
E: sales@data-storage.co.uk
W: www.data-storage.co.uk

## Our ancestors

Over the centuries the human race has tried to keep a history of events, discoveries and tragedies. Initially man carved etchings into stone or wood, then man discovered paper for almost 2000 years this was the way knowledge was passed on for future generations. Now in the 21st century a need to keep our digital history is becoming a huge headache for many organisations.

In the world we live in today we are constantly generating information in the form of video, sound, e-mails, texts, music, databases, images, office documents, PDF's, and many other varieties of data are created daily. Some of this information is kept for a day whilst much of this information needs to be kept for a month a year or more than 20 years. We keep this information for a number of reasons; it could be legislation, regulation, insurance, assurance, research, intellectual property, personal, historical, medical or many others.

## Human nature

I have been involved in the data storage industry for almost 30 years and during that time I have seen numerous technologies, fail to make the grade. The reason that they do not always succeed are that someone designs, develops, invents something that appears better, cheaper, faster, easier to use and more reliable. We are human and because of this we are seduced into purchasing these new revolutionary technologies. Unfortunately we do not always consider the long-term downsides of adopting these new products and it isn't until we come to upgrade, maintain and enhance this product that we realise it is not as revolutionary as first thought and we are locked in to a solution that cannot evolve with our day to day business demands.

Many organisations and individuals today are overwhelmed by the amount of information computers and people are generating. The cost of storing this information might initially seem to be affordable, but identifying and keeping this information for 20 or 50 years is proving to be a huge issue for business, even companies storing data for a short period of time up to 5 years are struggling.

## Data Compliance

Any archiving technology should be fully compliant with corporate or government legislation. A data compliant media should be 100% tamperproof, once information is written it cannot be altered in any way and this should be provable in court.

## Data archiving

Companies seem very reluctant to implement a tiered archiving solution whereby the most frequently accessed information is stored on the highest tier and the information that is infrequently accessed is stored on the lowest tier. The reluctance for organisations to archive data is based upon lack of knowledge and the fear of losing valuable information. The mindset of "keep it on RAID" arrays and backing it up every night is still the example we encounter daily as a business.

**80% of stored data is inactive after 60 days**

Numerous organisations have a huge on-going investment in purchasing data storage systems and this investment is increasing year on year as the demands to store more information increases. After 3-5 years this equipment is then replaced.

With the implementation of a Data Archive solution we can actually extend the life of this investment by moving the data to a secure active archive, thereby freeing up valuable disk space on

high performing storage solutions and slowing down the necessary and ongoing investment of more storage space giving a huge ROI benefit. An additional benefit with a tiered data archive is that you might be able to utilise your existing older storage systems to archive data.

As a consequence the size of online data volumes are spiralling out of control and storage management has become an ever-increasing challenge.

- Server performance and data access are diminishing
- Business legislation and user demands are requiring companies to increase disk space to alleviate the problem
- Data management puts high overheads on networks
- DR policies for recovery take longer as all data needs to be restored rather than the most essential
- The annual cost of managing this data over its lifetime is more than 5x the initial purchase price of the equipment
- Backup windows are diminishing

Buying bigger and faster storage systems is not solving the problem of data identification. How we classify our data is going to become more of an issue than people first thought. Now we know tools exist to move data by date, last modified, file size, extension, user, file type etc, but how many companies are using this technology?

An ideal archiving technology should be:

1. An open standard that is to say that it is not proprietary or owned by any one organisation.
2. Have a large installed base of machines that can read this vital information.
3. Should not be reliant upon proprietary software or hardware.
4. Can be stored in a normal office environment.
5. Can easily be duplicated.
6. Consumes minimal power reading or writing data.

Two types of technology immediately spring to mind one is CD the other DVD. These have both been around for 20 years and have a huge installed base and can be read today on a number of devices. They are cheap to produce, duplicate and consume no electricity. The issue with CD's and DVD's is their capacity is relatively small compares to today's disk drives.

## Tape

Many companies today consider tape to be an archiving technology, believe me it is not. Tape is and always will be media for restoring backups. My reasons for this are as follows;

1. A tape is a continuous spool of magnetic media, each layer is in contact with the next and overtime causes magnetic bleed which in turn causes data drop outs.
2. Tape needs to be stored in a humidity & temperature controlled environment, otherwise it starts to corrode.
3. Tape stretches with use.

Now all of the above is true but this isn't one of the reasons why we shouldn't use tape.

The reasons we shouldn't use tape as an archiving medium are as follows;

1. Tape formats are constantly changing and evolving. In the past we have had Exabyte, quarter inch cartridges, half inch tape, DLT, DAT, AIT, DC2000, LTO and many other formats that have appeared and disappeared with time. An example of this is the latest LTO-6 tape drives cannot read LTO-1 tapes created six years ago.
2. The software and operating systems to write these tapes with each version or release changes and eventually these tapes become unreadable. The software vendors that write the backup software also disappear.
3. Mechanical failure.
4. There are standards for writing tapes but very few companies actually use this and choose to write data in proprietary formats.
5. We also have to consider that some of these tapes use data de-duplication and encryption, if we lose the keys for either of these technologies the data is lost.
6. If we change our backup software from one vendor to another can you read an archive tape? The answer is probably not.

As I mentioned at the beginning, tape is a perfect technology for use in backups, it has a low-cost per gigabyte, is removable and reasonably robust.

## RAID Systems

RAID systems are not an archiving technology and can suffer from disk failure, controller failure, and mechanical failure, they are not very energy efficient and require constant monitoring to prevent data loss. A lot of archive information is stored on RAID systems which in 99.9% cases are protected using a RAID level. If we remove a single disk from the system the RAID fails unless a hot spare is available. If we advance 20 years what are the chances of having disks that can become part of that RAID set, absolutely zero. On top of this we have consumed copious amounts of energy and spent a small fortune on maintenance.

Another downside of RAID systems is that you cannot leave them in a Data Safe for protection against flood or fire and they are not very portable. Replicating RAID systems across WAN links to a DR site also adds to the cost of running a long term archive on RAID.

# Optical Storage

As previously mentioned optical storage has many benefits over the previous two technologies, the primary issue is capacity and installed base. This leaves two remaining technologies:

1. UDO – Ultra Density Optical
2. BD – Blu-ray

## Plasmon UDO

Plasmon in the early part of 21st century decided to develop a technology which they hoped become the de-facto standard for long term archiving. The issue they had was the cost to develop and maintain the development of the technology. On the surface UDO ticked many of the boxes required.

1. UDO-1 30GB (15GB per side) / UDO-2 60GB (30GB per side)
2. Housed in a ruggedised canister
3. 50+ year archive life
4. Random access to media
5. Fully data compliant – Phase change media once changed cannot revert back
6. True WORM – Write Once Read Many provides 100% data authenticity

Plasmon as a company went into administration in November 2008 due to the issues of developing a proprietary technology that had no OEM adopters. It also failed for three other reasons.

1. Very small installed base of machines with readers
2. Required software to read UDO media
3. Not written in a native format
4. No future roadmap beyond UDO-2

Plasmon UDO technology was purchased in 2009 by Alliance Storage Technologies and is now manufactured by them.

Another company that was developing holographic storage was InPhase Technology, but this also failed due to the huge investment costs for development to bring the product to market.

## Blu-ray

This leaves a single remaining technology Blu-ray which has many of the features of UDO but is supported by over 300 companies globally. A major advantage of Blu-ray media is that it is portable and easy to store in a Data Safe, so should a fire destroy the jukebox the media is still intact. Whilst Blu-ray is not sexy, particularly fast or has a huge capacity it is a technology that has evolved from the days of CD/DVD, is backward compatible and can read the predecessor's formats that make it a viable technology on which to place important digital information for future generations.

The main benefits of Blu-ray are:

1. 25GB, 50GB or 100GB written on a single sided Blu-ray disc
2. Super tough coating on the media to prevent scratches and damage
3. 50+ year archive life (shelf life)
4. Random access to media
5. Fully data compliant – Phase change media once changed cannot revert back
6. True WORM – Write Once Read Many provides 100% data authenticity
7. Huge installed base
8. Data can be written in UDF (Universal Data Format) which is a global open standard
9. Any machine with a Blu-ray drive can read the contents
10. Consumes very little energy unless required

## Archive Media

No matter what type of archiving technology you choose it is important to remember that the media is more important than the device. I know of a UK Police force that saved a few pence on non-branded DVD media, 2-5 years later when cases come to court the media is blank. If you do need to keep media for a considerable time always buy branded media from a known entity and ask for a certificate of conformance. These are sometimes difficult to obtain but should state that the media is certified for "x" years archive life. It might cost you a few pence more in the beginning, but it could save you thousands over the lifetime of the archive.

To ensure a long-term a long term archive you should ideally make three copies of the media and store it in three separate locations, if possible on three differing types of technology in the hope that in 50 years one of the technologies is still around to read the stored information.

## Technology refresh

With any technology you should perform the following:

- Periodically do a sample check on your archived media to ensure that the integrity of the media and data is still intact.
- Perform an analysis of current technologies and decide whether or not to keep or transfer your archive to a newer technology
- Ensure that the equipment is fully maintained and that you continue to purchase branded archive grade media
- Should you decide to change operating systems, always check compatibility with your archived data
- Ensure that the technology you are purchasing has a long track record and is not in the process of selling off products to please investors
- Do not de-dupe or encrypt archived data as ideally if you do need to restore the information in 20 years you want it in native format.
- If you are required to store data for 5+ years put it on optical it's far safer and will save you £'s in the long run.

## Summary

At the time of writing this document I know of no other technology under development that can provide the long-term archival storage requirements and installed base that Blu-ray currently enjoys.