



# The Role of Tape

*Why tape can be vital to an investigation.*

**Abstract:** This Whitepaper discusses the benefits of using backup and archive tapes as an evidential source in investigations, litigation and regulatory enquiries. The document outlines the difference between back-up and archive data, why tape is used and the challenges it can pose. It then discusses the benefits of using a non-native restoration solution and the reduction in cost that this delivers.

## The benefits of Non-Native restoration

## The Role of Tape

### **INDEX**

<b>Executive Summary</b> .....	3
<b>Why is tape used in an IT infrastructure?</b> .....	4
<b>How is information moved to tape?</b> .....	5
<b>What data does tape hold?</b> .....	6
<b>So why is tape such a good place to look for evidence?</b> .....	7
<b>There is another way</b> .....	9
<b>The cost benefits.</b> .....	10
<b>About eMag Solutions</b> .....	11

**eMag Solutions Limited  
2A Oaktree Court  
Mulberry Drive  
Cardiff Gate Business Park  
Cardiff  
CF23 8RS**

[www.emagsolutions.co.uk](http://www.emagsolutions.co.uk)



## Executive Summary

Information stored on various types of electronic storage is becoming central to virtually all types of corporate investigation be it litigation, regulatory or audit driven. The adoption of email as the standard communication and document distribution mechanism, coupled with its exponential growth, has focused attention on electronically stored information.

When conducting an investigation, where you look for that information can be critical to both the success and cost of the project. Accessing data from the 'live' computer systems is the most obvious place to start. However, where an investigation requires copies of emails and/or documents from the past – anything more than 3 months or more – then live computers cease to provide good returns.

Where older emails and documents are required, then the best – and often only – place to look is on backup and archive tapes. These tapes are routinely used by virtually all business to retain archive copies of their critical business information – often for many years. They provide complete and regular snapshots of a business, with all recorded communications and documents, over an extended timeframe.

Using conventional methods, tape has a reputation for being difficult and expensive to process and is often not considered as an evidential source. This can be especially so in litigation where the perceived cost is cited as being disproportionate to the value of the information likely to be uncovered.

However, there is an alternative way of accessing information stored on tape that can quickly and cost effectively produce the information you are looking for. The non-native system employed by eMag Solutions can rapidly process tapes, containing huge amounts of data, to deliver to you just the emails and documents relating to the people in whom you are interested, for the period you need and containing the words and phrases that suggest those emails and documents will be of use.

This document explains how and why tape is used in a business, what information it holds and why it can be crucial to an investigation. It then goes on to highlight the areas in which non-native restoration can deliver very considerable cost savings.

## Why is tape used in an IT infrastructure?

Ever since the introduction of the delete button there has been a need to keep copies of data stored on electronic systems. Today, the need to keep copies of critical business data is more important than ever, with companies generating ever increasing amounts of information every day. In particular, the adoption of email as the de-facto inter-personnel communications tool has led to a massive increase in the amount of documents being created and in the amount of electronic storage space needed to keep these.

There are two fundamental reasons that a business keeps copies of critical data. Firstly a business needs to know that if some unplanned event interrupts their normal processing, be it an accidental deletion of an important document, computer system crash, a component failure of something catastrophic like a fire, then they will quickly and efficiently be able to recommence operations using very recent copies of that critical electronic data. This need for business continuity or disaster recovery means that a business will routinely take very regular copies of their data for the specific purpose of using it to recover from interruption caused by an un-planned event.

This process of taking regular copies of data for business protection is backing-up. These tapes are kept for quite short periods and will be used to quickly restore data if something happens to damage or delete information from the computer system.

The other reason for taking copies of business data is to retain them, long-term, as archive copies for legal and regulatory purposes. This process is archiving data. Essentially what happens is that at the end of a set period – a week, month, quarter etc. an extra set of backup tapes is created that contains a full copy of the data being used that day. This is then stored away in an archive for possible future use.

In both of these cases, once the copy of the data is taken, the tapes are usually moved to a storage location away from the main area of operation. This could be a separate building or could be a different location completely. The reason for doing this in the case of backup data is to ensure that the copy is protected from damage if the main site is hit by a flood, fire or other invasive incident. In such a case the backup could quickly be retrieved and be used to get the business operational in a very short space of time. In the case of archive data, the tapes are moved to a separate site, both for protection and also because they can take up a considerable amount of space.

To retain either backup or archive data, a storage medium is required that is capable of holding high volumes of data, can be reused regularly over a long period, can easily be transported to move it away from the main place of business and is sufficiently low cost to allow multiple copies to be made. Magnetic tape meets all of these requirements providing a highly durable, portable, low cost bulk storage medium.

In summary, tape is routinely used to retain copies of all business documents and communications that are stored long term and away from the main place of business.

## How is information moved to tape?

To make a copy to tape of the business documents and information stored on computer essentially uses three components. The actual configuration and complexity of these systems can vary greatly depending on the scale of the company but essentially these components are a tape, tape drive and backup software.

The actual mechanics of transferring data via a tape drive to a tape varies little from the way that audio or video equipment records to a cassette. Tape passes a read/write head and a magnetic pattern is written to the tape that means the programme recorded can be replayed on demand. The only difference is that it is data that is electronically stored.

The backup software is used as an interface between the computer itself and the tape drive and copies the information from the computers disk to the tape in a form that makes it possible to read back – or restore – when required.

This process, with such straightforward roots, has been complicated in a way that only the IT industry can manage.

As the market for backup and restore products grew, many different manufacturers brought out new and different formats of tapes, drives and of backup software. As the volume of data that needed to be backed up also grew – especially as eMail gained a hold as a communication standard – then the pace of new product development and release accelerated. Companies needed – and were actively encouraged by their IT suppliers – to reduce the time spent backing up data (the “backup window”) to limit the time systems were unusable due to this process.

Since many of these products were produced by companies in direct competition with each other there was virtually no commonality between them – and often a single company would release software that had no ability to use the data created using earlier versions.

What all of this means is that there are millions of tapes stored in archives that hold data that is still relevant and still within retention periods that make it of interest to current litigation and other investigations. These tapes have been created using a huge number of different and incompatible systems. Furthermore this pool of tapes – and this problem - is still growing on a daily basis.

## **What data does tape hold?**

Backup and archive tapes typically hold a complete copy of all electronic data used by a company. Each different software system employs different ways of selecting and reducing the data written to tape but as a generalisation, tapes will hold a complete image of data held on the computer storage system at the point of backup.

This means that a tape will hold a complete copy of all emails stored before and up to the day the copy was made, along with all business documents, finance information, operational transactions as well as all of the computer programmes etc. that make the system run.

Where email data is stored, it will retain copies of emails that a user has marked for deletion but that have not been 'purged' by the systems administrator. This procedure is often only run on a monthly basis which means that there is often an on-going record of communications that the user believed removed some weeks before. (Tape will not however have a record of an email that is sent and then immediately deleted).

Since tapes are used to retain archive copies on a regular basis they will contain a snapshot of the business on a regular basis allowing full analysis of any particular period and detailed comparison between any two or more periods to enable things like document changes to be examined, email trends to be established or the identification of deletion patterns.

## **So why is tape such a good place to look for evidence?**

Tapes hold a complete copy of a computer system at the point a backup was made. This means that all information created and not deleted to that point will be present. Almost all investigations into electronically stored data involve a request to examine emails from one or more named people over a set period of time.

Tape is the ideal place to search for this information and, if the time period for which mail is required is more than a month or so ago, can be the only place worth looking.

Tape will hold a file that contains all the emails an organisation sends or receives. Furthermore, this file will also contain all emails that a user deletes up to the point an IT administrator decides to clean up the deleted emails – often only once a month. The real benefit of tape lies in the fact that even if the user has deleted the email and the IT administrator has cleaned up the deleted messages, then simply going back to the tape created immediately before this could well yield the information you need.

Thinking about the way that offences such as fraud or deception are carried out, they typically evolve over time rather than result from a carefully premeditated plan. A user will notice that some internal control is lacking or will see that they have inadvertently benefited from a situation and it is only at that point that they set out to exploit it to their advantage. From then on they may be careful and be wary of leaving any evidence of their actions but usually it is too late. EMail and other documents already created will expose their intent and serve to support a case against them.

Additionally, since the creation of tapes and their movement off-site is a back-office function often carried out away from the main place of business then users are often completely unaware of their existence. This means that they will often take only sufficient steps to remove evidential traces from their own systems, oblivious to the fact that daily copies are being retained and stored for years.

Where an investigation needs be covert, may involve a member of the IT staff or needs be carried out without being general knowledge then the fact that tape is routinely stored off-site at a third party vault means they can easily be collected and investigated without anyone outside of the investigation ever knowing.

Where current or recent email is required, convention dictates that an image of the 'live' email system be taken and interrogated and this is probably the correct and most effective option. However, if this is intrusive and causes too much disruption to the operation of the business – or the investigation needs be covert or low key - then as above, tape that has been stored for DR purposes presents an ideal alternative.

In the event that an investigation requires information other than eMail - or documents transmitted by email – such as financial data etc. then tape is equally suitable.

## **How do we get to this information using *conventional* methods?**

Tape is clearly a key, very valuable and often unique source of information. In litigation its importance was recognised in the 2006 changes to the Civil Procedures Rules, where the definition of *a document* was changed to recognise electronic communications, tape was specified as a storage medium that had to be considered in investigations.

However, tape has something of a reputation of being very difficult and expensive to handle which results in there being a very high cost associated with using the data off tapes. Indeed, it is the perceived cost of processing tapes that often cited in proportionality arguments against using this data.

There are many contributory factors to this notion that tapes are expensive. The primary reason is that exactly the same combination of backup software and tape drive that was used to initially create the tape has to be employed to perform a restore. Furthermore, to access much of the data – especially eMail - means that the whole environment from which the data was copied also has to be present.

The cost and logistic implication of this can be significant. It is extremely rare for an organisation to still have this legacy equipment – and little is still available to purchase. Much of the original computer system would also have to be rebuilt to mimic the way it was set up when the tape was created. If they do have all of the available equipment and software, then the skills required to use it are often long gone. Even if there is the equipment and skill set available to restore all of the data then the amount of spare storage required to hold the data can be enormous and very expensive – especially as it may not be required after that particular investigation.

The data itself would also pose significant challenges. Taking regular snapshots of the data and storing them away as archive does ensure that a complete record exists of the state of play at that point. However, when you then restore several of these archives – as you would if an investigation spanned a period of time – then you will often restore multiple copies of the same document. You would also restore many copies of completely irrelevant files and, where you restore eMail files you will also have several duplicate copies of many email. Long forgotten passwords may be required to access the emails or documents created by individual users.

The net result is that if a legal team wants, they can, by assuming a conventional systems approach, represent even a moderate restore from tape request to be extremely difficult, expensive and time-consuming delivering a vast amount of data for an equally costly review – easily winning the proportionality argument.

## **There is another way .....**

The key to cost effectively accessing this evidence-rich media and of removing proportionality as a counter-argument lies in adopting a wholly unconventional approach.

Rather than assuming that the only way of restoring from the tapes is to use the original – or native – environment, eMag Solutions have developed a Non-Native Restoration Engine that removes the need to employ any of the original hardware, software or computer system when restoring legacy data.

Over 20 years development has delivered a solution that can restore directly from tapes with no consideration for the tape type, history or backup systems used. Additionally, this restore engine can selectively restore just files that contain emails or the document types in which you are interested.

Additional features mean that the engine can recognise the dates on which tapes were created which can quickly identify which tapes and can then restore them without needing to know their sequence or any additional system information. Since the data is being restored outside of its original environment, the need to use any set passwords is also removed.

As files are restored from the tapes, the engine incorporates a sophisticated system that recognises if a particular file or email – with identical content- already exists and then removes it, retaining a record of its existence.

Only emails from people in whom the investigation are interested need be retained and these can then quickly be searched for specific words, phrases or combinations of both. Furthermore this communication record can be presented pictorially to show patterns and volumes of email traffic to quickly profile the way a user communicates and to show irregular or unusual contacts.

This means that you can receive for review just a DVD or disk with a single copy of each email or document which meets your very specific search criteria.

The time and cost that this saves can be massive, removing any financial argument that the other-side might use to prevent this superb information source being used.

## The cost benefits.

Using non-native means of restoring data from tape can significantly reduce the cost of accessing archive data both in terms of the time it can save and the resource and skill required to deliver the required information. The areas where the use of non-native solutions can avoid cost in areas including:

**Software & Licences:** The data being restored may have been created using software that has since been replaced or retired or may have been an older version of a product still in use. The user performing the restore may also opt to create a new system separate to the main 'live' IT environment to receive the data. In these cases, the user may have to purchase new software or additional software licences before this can be done.

**Hardware:** Similarly, the archive data may be on a type of tape that has been superseded and is no longer in use. In this instance the user would have to locate and purchase suitable drives.

If the decision is made to create a separate restore system, then additional server and systems may also have to be found or purchased.

**Storage:** There is likely to be a very large volume of data that has to be restored before any specific information can be found. It is highly unlikely that any company will have this amount of spare storage capacity available and will have to purchase new capacity. It is likely to be unused once the restore project is complete.

**Systems:** Using conventional methods, the environment in which the data was created will have to be recreated. This means that an appropriate (for example) Exchange or Notes environment will have to be created to receive the data. It is also probable that new software or utilities will have to be purchased to assist with things like the selection of mail for certain people, removal of duplicate mails and documents and the selection of mails that contain key words and phrases.

**Skills:** The restoration of archive data and the extraction of specific information for use in possible legal proceedings require a skill set not usually present in most corporate IT departments. Additionally, it is likely that the continual upgrade process that corporate companies engage in will have seen changes to the hardware, software and operating environments in use. All of these mean that additional skills will have to be acquired or brought in before a restore project can commence.

**Time:** When a request is made to access electronically stored information and produce selected emails or documents it is likely that the results will be needed within a defined timeframe. To identify the data sources, build a suitable restore system, recreate the original environment, restore the data, search through the results and identify the responsive data will take significantly longer than the time allowed. The cost to the client of that time and of missing deadlines will also be significant.



## About eMag Solutions

With a 40 year heritage in providing tape and data centric solutions, eMag is a global service provider. Core to all service delivery is an ability to restore from virtually any tape *without* needing to use the backup software or hardware that initially created that tape.

The main application area for this non-native restoration solution is in the field of regulatory or legal investigation, where the usual requirement is to produce historic email (sometimes 5yrs + ago) for named individuals from an extended period. Using the eMag non-native restoration system, this email can be restored directly from tape without needing to use any of the originating hardware, software or operating infrastructure.

The other key application area is in 'traditional' data service where data has to be converted or migrated between logical or physical platforms, a backup system has been changed and access is required to the data written in the retired format, legacy data cannot otherwise be read or in satisfying audit requirements around accessing or testing legacy data.

All services are available from one of 4 global production centres or on-site anywhere in the world using rapidly deployable mobile systems. This ability to quickly travel to any global location can be vital where prevailing data privacy laws mean that otherwise it would be difficult to export the data outside of the particular country borders.

Complimenting the extensive tape processing facilities is a complete suite of data services that are used to forensically capture information from any disk based storage device.