# An Overview of Backup Tape Challenges in a Litigation and Investigation Context

**By**
**John P. Collins, Vice President**
**The Ingersoll Firm**

Backup tapes store data which is copied from the various information technology systems used by an organization. Backup tapes have been used to store data for over 50 years; while other technologies are ascending in competition with backup tapes (for example, hard disk drives or "virtual tape libraries"), backup tapes remain the dominant media for backup and disaster recovery systems.

Backup and disaster recovery systems are employed by organizations to ensure they have up-to-date copies of their data available if some type of data loss occurs. Such data loss can range from the mundane (a user accidentally deleting an important file) to the catastrophic (a hurricane destroying an organization's primary data center).

It's helpful to consider some of the technical aspects of backup tape systems when contemplating identifying, preserving, and collecting relevant data from these systems. These aspects facilitate the efficient capture and storage of data on backup tapes; such aspects, however, significantly complicate discovery. These technical aspects of backup tape systems include:

- **Spanning:** this is the ability of a backup system to place a single backup job (such as the backup of a particular Exchange Server) on multiple tapes. In a normal restoration process, it is usually necessary to restore all tapes in a spanned set to retrieve the data.
- **Multistreaming/multiple data streams:** this is the ability of the backup system to take the backup for a single application or database—such as accounting data—and "split" the data into multiple streams which are simultaneously copied to two or more backup tapes. This increases the speed at which the backup job will complete.
- **Multiplexing:** refers to a process whereby data from multiple systems is combined onto a single backup tape. For example, data from an Exchange Server, accounting database, and a payroll system would all be contained on a single tape. As with multistreaming, multiplexing increases the speed at which the backup job will complete.

**Discovery challenges and issues presented by backup tape technical aspects:**

- Spanning impacts the discovery process by requiring multiple tapes to be restored, regardless of whether or not relevant data resides on the tapes. Spanning also complicates discovery by virtue of the challenges associated with keeping dozens of backup tapes in proper sequence; because backup tapes are rotated, or recycled, over and over, managing which tapes are in which sequence cab be problematic (especially in organizations that do little or no testing of backup tapes).
- Multistreaming/multiple data streams impacts discovery by causing the restoration of data to require multiple tapes to restore a single item of data. As with "spanned" tapes, managing multistreamed tapes is problematic. Often, tapes are taken out of tape rotation and placed in a location where they are forgotten; the more time that elapses, the more difficult it is to restore the tapes due to labeling issues, absence of personnel familiar with the tapes, and technology changes.
- Multiplexing results in the arbitrary combination of separate data streams onto a single backup tape; this means potentially unrelated data (both content and technology wise) is "woven" together (such as the multiple strands of fiber twisted together to make rope). The challenge with multiplexed backup tapes is preserving and retrieving JUST the relevant data and not the irrelevant data; because relevant and irrelevant data are

essentially bound together, both are preserved and retrieved, to the detriment of the producing party.

- For example, a multiplexed backup tape might contain four different data streams: Exchange Server e-mail; an accounting database; Word documents from a file server; and, a payroll database. When a piece of litigation arises, the producing party with the multiplexed backup tape may have to preserve ONLY the Exchange Server e-mail; the other data on the tape may be ignored. However, unless the party goes to the trouble of extracting only the Exchange Server e-mail (which frequently is NOT the case), all the data is preserved—thus, causing the non-relevant data to be preserved along with the relevant data, effectively making data that could have been destroyed available for future litigation. If the producing party somehow "unwinds" the combined data streams so that the relevant and non-relevant data is separated, then the relevant data can be preserved per the particular litigation, and the non-relevant data can be deleted per the organization's retention policies.

## Native vs. Non-native Backup Tape Restoration

### Native Restoration

The purpose of backing up data is to have copies of the data available in the event of a computer failure or disaster. Accordingly, the fundamental objectives of a backup tape system are:

1. Backing up data efficiently and without error.
2. Restoring data quickly and without error when the need arises.

These objectives do not take into consideration the need to restore data from backup tapes for litigation or investigation purposes.

The native restoration of data from backup tapes is designed to be accomplished essentially as the reverse of the backup process. The longstanding premise of tape backup systems is to copy the data and, if the data needs to actually be retrieved, to restore that data into the **identical environment from which it was originally created**. The backup software is designed to ensure precise, up-to-date copies are made; in this process, the backup software is creating a copy of the data that is the product of its environment. In a disaster recovery context, having a precise or mirror environment available into which to restore the data from backup tapes is presumed; it is also presumed that all the data must be restored, not just specific portions.

In litigation or an investigation, backup tapes may need to be restored at a time significantly removed from when they were created, meaning the identical or mirror environment is unlikely to still exist (the organization has probably upgraded software, changed some hardware, added new applications, etc.). An additional dynamic is that in a litigation or investigation, the entire body of data on a backup tape is probably not being sought, but rather, specific data related to the matter. As discussed earlier, backup tape technology, such as spanning, multistreaming, and multiplexing make identifying, preserving, collecting, and producing discrete, targeted data (as opposed to everything on the tapes) virtually impossible in a native restoration mode.

### Non-native Restoration

Non-native restoration employs a set of sophisticated technologies developed through the reverse engineering of backup tape hardware and software. Non-native restoration retrieves targeted data from backup tapes without having to replicate the environment in which the tapes were originally created; this means there is no need to re-create servers and hardware with the same operating specifications as existed at the time the backup tapes were created.

Non-native restoration bypasses the normal backup tape restoration process. Instead of using the methods and procedures of the backup software, non-native restoration employs methodologies which bypass the backup software and "rip" data off the backup tapes. In bypassing the backup software, non-native restoration avoids the need to provide a mirror environment from which the backups were created; thus, there is no need to recreate the Exchange Servers, database servers, or file servers which existed at the time the backups were made. Benefits of non-native restoration include:

- Reduces number of steps required to retrieve targeted data
- Does not require IT staff to build a restoration environment which costs time and money.
- Saves significant time as it does not require restoration of all the data to take place in order to retrieve targeted files; in some instances, targeted files can be identified on the tapes and only those files extracted

## Backup Tape Processing Options

1. **Native backup tape restoration.** In native restoration, IT professionals build a replica ("mirror") environment in which to restore all data residing on the backup tapes. A mirror environment requires software versions and hardware identical to the environment being re-created. Any discrepancy in software versions or hardware is likely to cause problems. Native restoration typically includes the following steps:
    a. Set up hardware: computers, tape drives, cabling
    b. Install software: application software, backup software, tape drive software
    c. Test to ensure environment is functional
    d. Restore tapes in proper sequence
    e. Retrieve targeted data using native software

3rd party vendors can provide native backup tape restoration services; some organizations attempt restoration using internal IT resources, usually with varying degrees of success. Also, using internal IT resources can be a drain on an organizations resources.

(While native restoration is time consuming and expensive, in some situations it may be the only option available.)

2. **Non-native backup tape restoration.** Specialized technologies and procedures are employed to "rip" data off backup tapes without building a mirror environment. In non-native restoration, data on the tapes is copied onto high performance computers, which run sophisticated processes over the data. There is no need to recreate the environment from which the tapes were created. The data will be extracted and placed onto alternate media, in the format requested.

The following non-native restoration services permit an organization to "see" what is on backup tapes without having to pay for full extraction. Often, just knowing what is on the backup tapes can be of value.

- Tape header reports
    - When tape was created
    - What backup software was used
- File listings
    - The files on the tape (.exe, .dll, .bat, etc., .pst, .ost, etc.).
    - Are e-mail files on the tape determined definitively
- E-mail database custodian report
    - Generate a list of the custodians whose mailboxes are contained in the e-mail database file contained on the tape
    - Includes the tape header and file listing reports
- Full extraction
    - This entails extracting all targeted data from the tape in its raw format, and includes the tape header reports, file listings, and e-mail database custodian report
    - With a full extraction, targeted data may be preserved onto alternate media, permitting the non-targeted data to be destroyed in accordance with the organization's retention policies

**Questions for obtaining price quote for tape extraction services.**

How many tapes?

What type of tapes?
- DLT
- LTO
- 4mm
- 8mm
- DTF
- DAT
- AIT

What backup software was used to create the tapes?
- NetBackup
- BackupExec
- CommVault
- CA ARCserve
- CA BrightStor
- Windows NT Backup
- Tivoli (IBM)
- EMC Data Manager
- HP OmniBack
- HP Data Protector

What are the contents of the tapes?
- Exchange Server (e-mail)
- File server (user documents)
- Database (Oracle, SQL, DB2)

Is the data on the tapes compressed?
What are you looking for? Be as specific as possible.
- E-mails with certain keywords?
- All of a particular custodian's e-mail?
- Word, Excel, or PowerPoint files?
- Specific databases?

What type of output do you seek?
- Native files (Excel spreadsheet = Excel spreadsheet; Word document = Word document VERSUS Excel spreadsheet = .PDF image or .TIFF image)