



What Judges Should Know About Discovery from Backup Tapes

Craig Ball

What Judges Should Know About Discovery from Backup Tapes By Craig Ball

© 2007

The electronic discovery wars rage in the mountains of e-mail and flatlands of spreadsheets, but nowhere is the battle so pitched as in the trenches of back up tapes, those vast electronic packing crates at the heart of front page cases like *Zubulake v. UBS Warburg* and *Coleman (Parent) Holdings v. Morgan Stanley*.

Why are backup tapes such troublemakers?

Ideally, the contents of a backup system would be entirely cumulative of the active “online” data on the servers, workstations and laptops that make up a network. But because businesses entrust the power to destroy data to every computer user—including those motivated to make evidence disappear—and because companies configure systems to purge electronically stored information as part of their records “retention” plans, backup tapes may be the only evidence containers beyond the reach of those who fail to preserve evidence and those with an incentive to destroy or fabricate it. Going back as far as Col. Oliver North’s deletion of e-mail subject to subpoena in the Iran-Contra affair, it’s long been the backup systems that ride to truth’s rescue with “smoking gun” evidence.

Backup tapes can also be the target of aimless fishing expeditions mounted without regard for the cost and burden of exploring tapes, or tape may be targeted prematurely, before more accessible data sources have been exhausted.

All of this becomes the court's problem when discoverable information may reside on backup tape and:

1. A party obliged to preserve ESI overwrites or discards tapes;
2. A party seeks to be released from a litigation hold on backup tapes;
3. Restoration and review of backup tape will engender significant delays; or,
4. The burden or cost of tape restoration is unreasonable given the circumstances of the case.

Grappling with Backup Tapes

Backup tapes are copies of data on a computer obtained for *disaster recovery*, i.e., picking up the pieces of a damaged or corrupted system. Some call backups "snapshots" of data and, like a photo, backup tapes capture only what's in focus. To save time and space, backups typically ignore commercial software programs that can be reinstalled in the event of disaster, so *full backups* focus on all *user created* data and *incremental backups* grab just what's been created or changed since the last full or incremental backup. Together, they put Humpty-Dumpty back together again in a process called *tape restoration*.

Jargon Watch
disaster recovery
full backup
incremental backup
tape restoration
tape rotation
legacy tapes
serial access
vertical deduplication
horizontal deduplication

Tape is cheap, durable and portable, the last important because backups need to be stored away from the systems at risk. But, tape is also slow and cumbersome--foibles forgiven because it's so rarely needed for restoration. Back up systems have but one legitimate purpose, being the retention of data required to get a business information system "back up" on its feet in the event of disaster. A business only needs disaster recovery data for a brief interval since no business wants to replicate its systems as they were six months or even six weeks before a crash. As the only backup tapes that matter are the last complete set before the river rose, *in theory*, older tapes are supposed to be recycled by overwriting them in a practice called "*tape rotation*."

But, as theory and practice are rarely on speaking terms, companies may keep backup tapes long past their usefulness for disaster recovery--often *years* past, and even beyond the companies' ability to access tapes created with obsolete software or hardware. These *legacy tapes* are business records--sometimes the last surviving copy--but afforded little in the way of *records management*. Even businesses that overwrite tapes every two weeks replace their tape sets from time to time as faster, bigger options hit the market. The old tapes are often set aside and forgotten in offsite storage or a box in the corner of the computer room.

Like the DeLorean in "Back to the Future," legacy tapes allow you to travel back in time. It doesn't take 1.2 million gigawatts of electricity, just lots of cabbage.

Why is Tape So Slow?

Actually, tape is pretty remarkable technology that's seen great leaps in speed and capacity.

Still, there are those pesky laws of physics.

Tape is *serially accessed media*, meaning you must plow through its contents to get to what you're seeking. It's like trying to find the start of a show on a VCR tape: you fast forward or rewind to get there. What's more, information on tape may be recorded in a serpentine fashion, like a mountain switchback, so the tape drive must shuttle back and forth through the entire tape repeatedly to get to the data. As this is a mechanical process, it's occurs at a glacial pace relative to the speed with which computer circuits or even hard drives move data.

Although newer backup tape technologies build in some indexing features, older systems are limited in their ability to find and extract particular files. Further, recalling that backup is an incremental process, reconstructing reliable data sets may require data from multiple tapes to be combined. Add to the mix the fact that as hard drive capacities have exploded, tape must store more and more information to keep pace. Gains in performance are offset by growth in volume.

But remember, tape is cheap, durable and portable. As the cost of hard drives and other media plummets, tape will go the way of the floppy disk; but, it's with us for some years yet.

Deduplication

Parties rarely fight over restoration of a single backup. Companies that archive backup tapes may retain years of tapes, numbering in the hundreds or thousands. Because each full backup is a snapshot of a computer system at the time it's created, there is a substantial overlap between the backups. An e-mail in a user's Sent Items mailbox may be there for months or years, so every backup replicates that e-mail, and restoration of every backup adds an identical copy to the material to be reviewed. Restoration of a year of monthly backups would generate 12 copies of the same message, thereby wasting reviewers' time, increasing cost and posing a risk of inconsistent treatment of identical evidence (as occurs when one reviewer flags a message as privileged but another decides it's not).

Accordingly, an essential element of backup tape restoration is deduplication; that is, using computers to identify and cull identical electronically stored information before review. Deduplicating within a single custodian's mailboxes and documents is called "*vertical deduplication*," and it's a straightforward process. However, corporate backup tapes aren't geared to single users. Instead, business' backup tapes hold messages and documents for multiple custodians, many of whom store identical messages and documents. Restoration of backup tapes generates duplicates within individual accounts (vertically) and across multiple users (horizontally). Deduplication of messages and documents across multiple custodians is called (not surprisingly) "*horizontal deduplication*."

Horizontal deduplication significantly reduces the volume of information to be reviewed and minimizes the potential for inconsistent characterization of identical items; however, it can make it impossible to get an accurate picture of an individual custodian's data collection, since many constituent items may be absent, eliminated after being identified as identical to another user's items.

Tips on Tapes

Tape ≠ Inaccessible

Tape, like paper, is just a medium to store information. Knowing that information is on paper tells you nothing about the complexity of the content. Is it, "See Spot run" or particle physics? Is it in English or Swahili?

Litigants may mistakenly speak of backup tape as "inaccessible," but unless physically damaged or written in some obscure format, tape is not inaccessible. What they *mean* to argue is that the expenditure of resources required to access the contents is unreasonable.

Accessing some backup systems requires no more than popping in a tape and searching its contents. But, backup systems of large companies with multiple business units are enormously complex and may lack straightforward correspondence between individuals and data. Every system is different.

Consequently, there is no rule-of-thumb for accessibility of backup tapes. When challenged, it's the responsibility of the party claiming inaccessibility to bring forward proof of unreasonable burden or cost.

Proof

Proof must consist of more than lawyer gesticulation and stentorian protest. Demand testimony of, e.g., IT personnel, outside vendors or qualified experts recounting the actual cost and time to complete a restoration of information on tape to a more accessible form, but excluding (or at least segregating) the cost of search, filtering or attorney review, these being costs ordinarily borne by a responding party searching accessible ESI.

The court may want to inquire:

- Does the responding party routinely restore backup tapes to, e.g., insure the system is functioning properly or as a service to those who have mistakenly deleted files?
- Have any of the backup tapes at issue been restored in other circumstances and thus accessible as information in other cases or held by third parties?
- Does the responding party have the system capacity and in house expertise to restore the data? *Not everyone has the idle system resources or personnel required to temporarily restore a prior version of the data alongside the current version.*
- Can responsive data be searched and extracted without wholesale restoration of the tapes? *Emerging software and tape technologies sometimes make this feasible.*
- Have you compared your projected in-house restoration cost against the services of so-called “tape houses” equipped to process large numbers of tapes at competitive prices? *“Do it yourself” is not always cheaper.*

A Threshold Question

The oft-overlooked threshold question is, “What is the chance the tapes contain evidence?” If backup tapes potentially hold responsive information, but a responding party declines to search or produce the contents, the responding party must identify the tapes and, as feasible, “provide enough detail to enable the requesting party to evaluate the burdens and costs of providing the discovery and the likelihood of finding responsive information” on the tape. *Committee Notes to FRCP Rule 26(b)(2)(B).*

How can a party gauge the likelihood of finding responsive information on tapes without searching them?

The answer lies in recognizing that backup tapes don’t exist in a vacuum but as part of an information *system*. A properly managed system incorporates labeling, logging and tracking of tapes, permitting reliable judgments to be made about what’s on particular tapes insofar as tying contents to business units, custodians, machines, data sets and intervals. If the responding party has so poorly managed backup tapes that nothing can be apprehended about their contents, the court must decide whether it’s fair to deny access to the evidence or shift costs. In such case, the Court may instruct the responding party to generate an index of contents but defer requiring a full restoration and review pending examination of the index.

First Pick the Low Hanging Fruit

Too often, parties battle about backup tapes before they’ve looked to see what’s in easily accessible sources and material already produced. Be sure that the parties have exhausted accessible sources for the information. Has the producing party searched the contents of all servers, desktops, laptops, external hard drives, handheld devices and removable media like CDs and floppies that may hold the information also stored on tape? Restoring a few tapes can start to look attractive to a party who hasn’t thoroughly explored active data repositories. Searching for information costs money, and it may be preferable to look in one hard-to-access place than a hundred easy ones.

Imposing Conditions on Discovery from Tapes

If the court determines that backup tapes are reasonably accessible, then the responding party is obliged to search them and produce responsive information at its own cost, subject to the FRCP Rule 26(b)(2)(C) limitations that apply to all discovery. FRCP Rule 26(b)(2)(B). But, if the court determines that the tapes are not reasonably accessible based on cost or burden, then the requesting party must satisfy the court that there is good cause to compel production from the tapes. *Id.* If the court then opts to order discovery from the tapes, the court may wish to specify conditions for the discovery to ameliorate hardship and guard against overreaching. Such conditions might include:

1. Sampling to explore relevance

Sampling backup tapes is like drilling for oil. First, you identify the best prospects. If a prospect yields several dry holes, you move on. But, if it starts producing, you keep on drilling.

Sampling backup tapes entails selecting a part of the tape collection deemed most likely to yield responsive information and restoring and searching only those selections before deciding whether to restore more tapes. The size of the sample hinges on many variables, among them the breadth and organization of the tape collection, relevant dates, fact issues, business units and custodians, resources of the parties and the amount in controversy. The parties should agree on a sample size. If they cannot, the court may wish to seek guidance from a knowledgeable neutral or simply apply its sense of fairness. Recognizing that a backup snapshot often consists of more than a single tape, the court should insure that each sample is complete for a selected date; that is, the number of tapes shouldn't be arbitrary but should fairly account for the totality of information captured in a single backup interval.

As important as the sample size is the question of who gets to select the backup sets for particular dates, machines, business units or custodians that will comprise the sample? Here, the best approach is to either allow the requesting party to specify the goal of the examination and require the producing party to tailor the selection to best meet that goal or require the producing party to share sufficient information and technical assistance to allow the requesting party to reliably gauge what might be on the tape (by date, business unit, users or whatever information can be ascertained), then let the requesting party choose the "slices" of data to restore and search.

2. Shifting the cost of making information on tape accessible for search and review

Cost shifting is a magical tool. It has the power to transform parties into more reasonable, efficient and cooperative creatures or to slam the courthouse door to persons of modest means with meritorious claims and defenses. If the court determines that backup tapes may contain responsive information, the court may order the requesting party to bear the reasonable cost of converting the contents of those backup tapes to more accessible forms, e.g., the cost to have a tape conversion service provider extract the compressed data to accessible formats on external hard drives. Once the information is made reasonably accessible, the cost to review and produce from the accessible sources remains the responsibility of the producing party.

This is frequently an equitable approach, but it should be used with care. A party's willingness to bear the cost of restoration is insufficient to justify discovery from sources unlikely to contain responsive information, and it may nonetheless impose an unreasonable burden in terms of the added volume of data for review. The well-heeled shouldn't get broader access to an opponent's ESI or unfairly increase an opponent's review burden just because they can afford to be curious.

3. Permitting an inspection of contents

Whatever happened to making information available for inspection as it was kept in the usual course of business? Backup tapes are the modern counterpart to banker's boxes in the warehouse. The court

may wish to explore the feasibility of affording a requesting party access to duplicates of backup tapes in order to permit them to inspect contents, with appropriate safeguards for privilege, privacy and trade secret considerations. The producing party may need to offer technical assistance to insure that the requesting party can access the information on tape. This may entail, *inter alia*, sharing passwords required to decrypt locked data and details concerning the computing environment and backup software and hardware. It may be necessary for the producing party to afford access to technical manuals or versions of backup software no longer commercially available. *Compel cooperation.*

Restoration of Backup Tapes in Mitigation of Spoliation

A circumstance where production of responsive information from back up tapes may be unconditionally ordered notwithstanding cost or burden is where the responding party failed in its duty to preserve information in accessible sources. This can occur when, by guile or negligence, a party deletes, discards or corrupts information that should have been subject to a litigation hold directive or when a party fails to discontinue routine migration of information from sources easily accessed to less accessible formats. Certainly, an organization need not discontinue its routine archival of information, but neither should that migration operate to deny an opponent access to information that the producing party was bound to preserve for use in an anticipated suit or claim.

Conclusion

Backup tapes epitomize the cross purposes of information technology and litigation. Compelling a large organization to interrupt its tape rotation, set aside backup tapes and purchase a fresh set can carry a princely price tag; but if the tapes aren't preserved, deleted data may be gone forever. Must a litigant forego essential evidence or pay more to secure it than the amount in controversy? *In a world where virtually no electronic evidence disappears, but only gets harder to reach, how much is enough?* These are a few of the Hobson's choices of e-discovery, but they are made easier when the court understands the technical challenges and can help fashion practical, equitable solutions.