

Market Report

Tape Remediation

Providing Measurable Capital and Operational Cost Reduction Options

By Brian Babineau

June 2010

Contents

Introduction	3
The State of Tape	3
A Wide Range of Challenges	4
It's Not That Cheap	4
Electronic Discovery Creates Unpredictable Costs	5
Decentralized Electronic Records Management	5
Tape Remediation Delivers Insight and Action.....	6
Knowledge is the First Step	6
Take More Discrete, Beneficial Actions.....	6
Finding the Right Solution	7
Prioritizing a Project.....	7
The Bigger Truth	8

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188.

Introduction

For most organizations, backup tapes represent an insurance policy for a critical asset: information. And because no one wants to operate without some level of insurance, tape is usually one of the first things funded in the IT budget—with few questions. If someone did examine this line item, they would quickly realize that there are significant expenses associated with managing and storing legacy data on backup tapes, including:

- Off-site service provider storage and transport fees to store, and if needed retrieve, old tapes.
- Maintenance and support of legacy tape infrastructures including systems and software needed “just in case” an old tape has to be restored for compliance reporting or electronic discovery purposes.
- Specialized consultant fees needed to restore and recover data from old tapes in response to electronic discovery requests.
- Opportunity costs associated with IT resources needed to restore and recover data from old tapes in response to electronic discovery requests if external consultants are not used.

Because IT departments rarely know what content is actually stored on a tape, the popular offline backup media can be a legal liability. There may be the proverbial “smoking gun”—an e-mail, contract, etc—stored on a backup tape that, if it is uncovered and produced during discovery, could substantially increase an organization’s legal exposure. At a minimum, corporate counsel would be better served if they knew such content existed so they can better prepare for a matter (i.e., settle before the file is shared with an opposing party or regulator). In an ideal situation, most organizations would move potentially risky content to online records managements systems where it can be expired according to established policies, eliminating the liability altogether.

Tape serves a useful purpose in today’s IT environments, but older media can be a financial and operational resource drain if it is not managed appropriately. The only way management of legacy tape data is if IT better understands the data on the tapes so they can take appropriate action. An investment in tape remediation will provide the appropriate level of insight so IT can:

- Delete duplicate data and reclaim media capacity.
- Copy business records and other potential litigious or regulated data into online repositories.
- Make aged data accessible without legacy versions of backup software and hardware.

From an operational perspective, these actions help reduce and consolidate tapes, lowering service provider storage fees. Business process improvement is derived from IT staff being able to quickly respond to and, more importantly, successfully satisfy electronic discovery requests without distracting them from normal operations or calling in specialized consultants. Additionally, electronic records management processes to address compliance and governance requirements are centralized as historical data (from tape) can be leveraged by more recent business records in online repositories. This paper helps detail these cost savings and risk mitigation opportunities so companies can prioritize a tape remediation initiative.

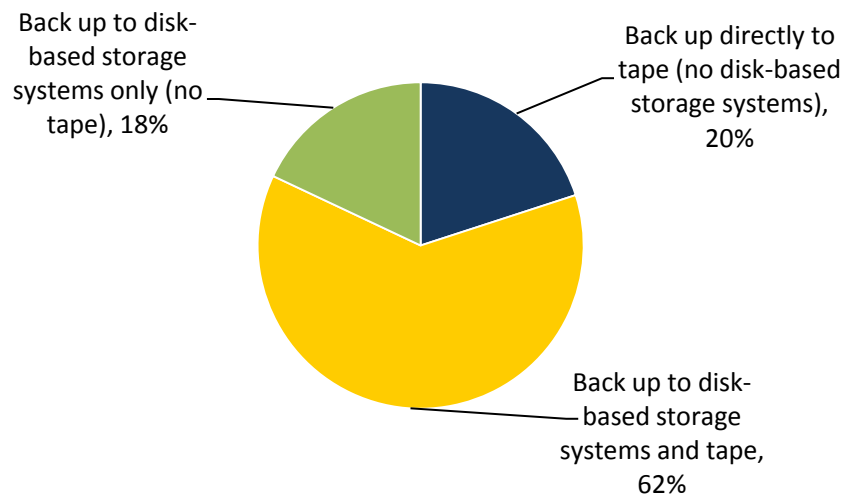
The State of Tape

Tape is the predominant storage media used for data protection due to its portability and, from an acquisition cost perspective, its price. What most do not realize is that, despite the uptick in investments in disk-based backup solutions, tape is still a staple in this area. According to recent ESG research, 82% of organizations still use tape to support all or a portion of onsite backup processes (see Figure 2).¹ This represents only a 5% decrease from a similar survey that ran in 2008.

¹ Source: ESG Research Report, [2010 Data Center Spending Intentions](#), January 2010.

Figure 1. Tape is Still a Part of Most Organizations' Onsite Backup Processes

Which of the following best describes your organization's on-site data backup process? (Percent of respondents, N=441)



Source: Enterprise Strategy Group, 2010.

One of the more telling statistics taken from this research is that organizations expect that the combined capacity of onsite and offsite tape will not change over the next two years. Yes, there will be a shift as more companies use disk for onsite backup processes, but offsite tape capacity is expected to increase. This is largely due to primary data growth, representing more data that needs to be protected. Despite predictions of its death, tape's demise as a protector of newly generated information is far from imminent. It still remains the most predominant repository for historical information—specifically vital business records, some of which could be a liability if requested as part of a regulatory or legal matter.

A Wide Range of Challenges

It's Not That Cheap

A typical tape budget includes obvious capital investments such as media to accommodate data growth and extended retention policies, and library replacements as warranties and leases expires. These expenditures can rarely be avoided because they represent the corporate data insurance policy. The operational expenses associated with tape are less transparent: staff is needed to manage tape media, rotate tapes, and reclaim unused capacity. Someone needs to monitor tape backups and reset them when, all too often, they fail. Tape systems, due to the robotics involved, usually consume a disproportionate amount of data center space, driving up indirect, but measurable, operating expenses associated with data protection processes.

Most large companies also pay service providers for offsite vaulting, including "recall fees" when tapes need to be restored for recovery purposes. Frequent or large scale recoveries involving many tapes may not be planned for and consume additional financial resources that may exist in the budget. Also related to recovery are peripheral costs associated with older libraries. In cases where data exists on legacy formats, companies are often forced to keep multiple versions of backup software and older hardware to restore data for recovery. In fact, organizations with several years/decades of data on tape usually have to keep legacy infrastructure around, adding more cost and complexity to data center operations.

Electronic Discovery Creates Unpredictable Costs

It is common for legal and regulatory matters to involve legacy information, often requiring electronic data from several years in the past (as a current example, see the high profile case involving Goldman Sachs in which e-mails from 2006 are being subpoenaed today). Much of this data often resides on tape—and its discovery can be long, painful, and costly.

At the front end of the electronic discovery process (identification), IT is often forced to fly blind as records indicating which applications reside on which tapes might exist, but visibility into the data itself will not. The only option is to restore all tapes that might contain responsive information, often after recalling the tapes from a service provider. Only after all the data and applications have been restored can IT identify relevant information.

To comply with the legal duty to preserve, all tapes relevant to a case must be removed from normal rotation to prevent spoliation. This is not easy when the volume of tapes with potentially relevant information is substantial. It is often very difficult to track all the tapes in normal operations—any mistakes made with media involved as part of regulatory or legal matter can limit the organization's ability to defend itself or pursue action. Of course, preservation is a moot point if IT fails to initially identify all relevant tapes.

With so much uncertainty in the identification phase, there is a very real risk that important information will be missed. Failure to produce all information to an opposing party can create the appearance that something is being hidden intentionally, which can contribute greatly to a negative judgment and possible legal sanctions. Organizations that do not document the tape recovery, restoration, and search processes as well as the subsequent preservation of that information will be very hard pressed to prove evidentiary chain of custody. This documentation is needed to ensure that the data has not been tampered with or deleted accidentally/maliciously from the time an initial discovery and preservation notice is received. Failure to execute preservation or properly document chain of custody processes can cast doubt on the integrity of evidence, adding to the risk of an adverse outcome in the case.

Electronic discovery requests have become so frequent, and the process so cumbersome, that IT shops sometimes outsource the operation to forensic specialists who can charge anywhere from \$450 to \$2,000 per tape. These expenses pale in comparison to the fines that have been levied and settlements reached because companies cannot properly manage electronic discovery—especially matters involving legacy data stored on tape.

Decentralized Electronic Records Management

The challenges associated with electronic discovery are reflective of tape's inaccessibility. Corporate records are locked away on tapes and stored offline, frequently off site, making it difficult to find anything quickly. Employees cannot access this information for business reference or planning purposes. And, as companies roll out electronic records management programs supported by document management systems and archive solutions, historical data stored on tape remains separate—only newly created records are saved in these centralized, more accessible repositories. Records managers, compliance officers, attorneys, and IT have to search all repositories and tape to locate all relevant records when requested. One could even argue that investments in electronic records management solutions go underutilized because historical data is excluded.

Tape Remediation Delivers Insight and Action

Knowledge is the First Step

The foundational issue of many tape costs and challenges is that IT does not know what exactly is on all the tapes. To date, backup software and service provider inventory management systems provide high-level information such as what systems were backed up to what tapes and where those tapes are physically located; they cannot track the individual content within files and messages stored on the tapes.

The traditional method of obtaining knowledge about data on tape is to restore terabytes of data to its native format using the backup software which created the copy in the first place. Once the restore is complete, the data can then be indexed and searched. This cumbersome, time consuming method can now be replaced by technology which can index the data directly from tape, skipping the entire restore process and avoiding the use of backup software. The index of the tape data includes the tape's property information (when it was created, what policy group it is a part of, etc.); file- and message-level details such as text content, file owner, etc.; and an indication of whether a particular set of backup files already exist somewhere else, a common occurrence where typical backup procedures call for multiple incremental and full copies. Using direct indexing of tapes, and eliminating the restore process, makes access to historical data affordable and efficient.

Take More Discrete, Beneficial Actions

Armed with this newfound, detailed knowledge of their tape environment, IT can take intelligent, informed actions to optimize tape operations including:

- **Deletion of redundant data on tape.** Most of a company's tape backups and archives were generated with procedures that create multiple copies of the same data. Companies should use the knowledge gained from remediation to identify and eliminate redundant copies of data on backup tapes that has no business value. As an example, if an incremental backup exists from the third week in April as well as a full month end copy from April, the incremental could be deleted because all of the data is captured in the full. Tape capacity can be further reduced by deleting irrelevant system files—a practice known as “deNisting.”

It should be noted that deleting any corporate data should consistently follow corporate information retention and disposition policies. As an example, a company may agree that, unless otherwise needed to satisfy a current legal hold obligation, any daily incremental backups or weekly full backups can be deleted so long as a monthly full backup exists.

- **Making more high risk data accessible.** To streamline responses to anticipated discovery requests, data likely to be responsive in legal matters archived on tape can be ingested into more accessible, manageable systems such as a disk-based archives or document management systems that contain other pertinent business records. Consolidation of legally sensitive information from inaccessible tape formats into a more “active” platform centralizes information for easier identification and preservation, and allows the application of records management policies to the data, including content expiration policies. It also helps companies avoid repeatedly searching for the same data on tapes, which often occurs when there are several inquiries related to the same individual (an executive that committed fraud) or the same subject matter (stock option backdating).
- **Eliminating expensive legacy infrastructure.** Companies incur substantial costs keeping and maintaining legacy backup infrastructure solely to support the restoration of data on old tapes in outdated formats. IT can identify which sets of legacy data still have enough business value to warrant saving and any data that exists in outdated formats can be ingested into more modern formats that comply with current data management standards. As a result, IT will finally be free to retire old legacy gear and software that was kept around (often unused) for years “just in case.”

Even if companies choose not to take aggressive action in their tape environments, direct indexing provides basic search capabilities that can aid in electronic discovery or even simple recovery activities. After indexing tapes, IT can input search criteria from a discovery or recovery request and determine what tape(s) need to be recalled (if they are off site) and restored. This eliminates a costly guessing game and shortens the time it takes to complete these tasks. Also, in discovery situations specifically, if certain search terms outlined in the inquiry return several thousand tapes, IT can advise legal of the cost burden of extracting all this information so they can potentially narrow down the scope. In contrast, if terms within the original request do not register on any tapes, IT doesn't have to complete the retrieval processes.

Finding the Right Solution

Indexing a tape environment is not an easy task given the amount of information that exists on any given tape, the number of media formats that exist across the typical enterprise (LTO, DLT, etc.), and variations in backup software required for restore operations ([IBM TSM](#), [EMC Legato](#), [Symantec NetBackup](#) and [Backup Exec](#), [CommVault](#), etc.). The easiest way to get the most out of any tape remediation project is to ensure the supporting solution performs direct indexing of the common permutations of backup software and tape media—both current and historical configurations. It is also important for the indexing solution to scale up to process many tapes and to provide a central management point that enables a consolidated view across the entire environment. Doing so expands opportunities to take more action with the information, including identifying and deleting duplicates.

To provide the benefits of migrating potentially more active information from legacy tapes to modern, more manageable formats, the remediation solution should enable the selection of and extraction of specific files and data from tape to be ingested into the new environment. Lacking this capability would require a full restore of all data from tape and then copying it to an online archive or other repository of record. In essence, IT should be able to copy and manage data directly from tape without having to restore it. With typically a small amount of data on tape being of interest, it is far more efficient to extract this small data set and leave the useless data behind versus traditional restore methods that require restoration of the entire tape before you can determine the contents.

If a solution can extract data from tape, it is critical that it do so with the metadata intact. This provides assurance that the actual data has not been changed even though it was extracted from tape to another source—a critical component to proving the integrity of the information for chain of custody purposes if the information is ever produced as the end of a discovery process.

Prioritizing a Project

The benefits and ROI of tape remediation are easy to see and should be relatively easy to measure. Among the key tangible metrics, companies should consider:

- **IT budget for tape media and systems.** Remediation allows the identification and deletion of redundant data that provides no business value, the reclamation of significant storage capacity, and the reduction of new tape media purchases. ESG continues to advise organizations that spend hundreds of thousands of dollars on new tapes every year to cut half of this expense—it would provide material cost savings.
- **Offsite storage fees.** Remediation reduces the number of tapes that must be shipped and stored off site, reducing service provider costs and “special charges” like recall fees. Service providers usually charge between \$5 and \$10 annually per tape—it may not seem like much, but these charges rarely go down because companies are always backing up more data and never deleting it. Tape remediation can help address the latter issue in line with company policies.
- **Legacy infrastructure retirement.** Only IT knows the maintenance fees for hardware and software associated with legacy backup and tape infrastructure. IT also should consider any specialized data center requirements or skill sets needed to run this outdated infrastructure. Tape remediation offers companies a chance to rid themselves of these expenses while keeping the information stored within the environment searchable and accessible if needed.

- **Tape-specific electronic discovery fees.** Every time IT has to spend resources chasing data on tapes for electronic discovery, there is a real opportunity cost. Help desks, system administration, and development efforts are left understaffed so resources can be diverted to support legal and compliance requests. Alternatively, IT simply outsources tape-related electronic discovery, which carries substantial expense. As suggested, remediation can drastically reduce the time and resources required to support tape-dependent discovery requests, minimizing the opportunity cost or the need to use external specialists.

Eliminating incremental tapes can immediately reduce the volume of tapes in storage by up to 60% which manifests in lower tape budgets and reduction in service provider fees. When it comes to intangible metrics related to tape remediation benefits, one only has to look at the risks associated with mismanaging electronic discovery. There are numerous opinions and sanctions issued by courts regarding companies' inability to properly preserve information. Rather than citing all of these here, ESG recommends reviewing a summary² of these compiled by Gibson, Dunn & Crutcher, a well respected global law firm known for its detailed litigation practice including a focus on electronic discovery.

The Bigger Truth

IT departments know that tape is not dying nearly as quickly as outsiders say it is, but they also rarely know how much tape really costs them. Many times, it is not the actual expense that causes problems—it is the lack of predictability and the risk associated with some of the costs. These issues frequently manifest in electronic discovery processes where IT and legal must figure out a way to get the job done.

Rather than trying to blindly delete or restore tapes as part of an effort to move data to an online archive, companies should take a bigger picture approach. They should first find out what data is on tapes and then take more precise management actions. When supported by direct indexing and discovery solution, these tactical actions, along with many others, can be undertaken as part of an overall tape remediation project. Such an initiative does not have to be a short-term project; it can be a phased approach where companies address bigger, more costly issues first and take on others as needed. Overall, there are plenty of achievable, measurable benefits that can be gained by IT, legal, compliance, and other departments prioritizing tape remediation as a business, not just a technology, initiative.

Implement Tape Remediation with Index Engines

Historically it was far too expensive and difficult to understand the content of old backup tapes. All data would need to be restored first, and then analyzed in order to determine what should be kept to support litigation and what could be purged. Restoring thousands or tens of thousands of tapes was out of the question. It would take too much time, money and legacy infrastructure.

The problem has been solved by eliminating the need for expensive and time consuming tape restoration and applying a more intelligent approach. Index Engines technology scans tapes, creating a searchable index of the content. Full text and metadata search makes finding relevant content fast and easy. Once content is discovered it can be extracted without the need for the original backup software. This allows you to only deal with relevant files (less than 1% of the tape content) and not the bulk of useless content. Index Engines intelligent tape discovery solution has made tape remediation an achievable project.

Index Engines' Tape Assessment Program outlines the clear return that any organization will receive by intelligently remediating old backup tapes. This program leverages the knowledge of our discovery experts and delivers a customized plan for your organization's implementation of new technology for tape remediation. This assessment outlines the timeline and process for the project and calculates the ROI that your organization will realize.

² Source: Gibson, Dunn & Crutcher, [2009 Year-End Electronic Discovery and Information Law Update](#), January 2009.

Tape Assessment Program – How It Works

- **Expertise.** Meet with Index Engines technical experts to review the process.
- **Customized.** Provide us with information unique to your environment, and Index Engines will analyze your environment.
- **Fast.** Index Engines will provide a Tape Assessment Proposal to you within 10 days.
- **Concise.** This plan will review how to implement a cost effective and efficient tape remediation project.
- **Economical.** Detailed ROI calculations will accompany this plan to clearly illustrate the money your organization will save, alongside reducing the risk unknown data residing on tape.
- **Start Today.** Contact Index Engines today to arrange for a tape assessment and understand how you can get control of your legacy tapes. www.indexengines.com, info@indexengines.com

² Source: Gibson, Dunn & Crutcher, [2009 Year-End Electronic Discovery and Information Law Update](#), January 2009.



Enterprise Strategy Group | **Getting to the bigger truth.**