



E-Discovery for Backup Tapes: How Technology Is Easing the Burden

by Jeffery C. Fehrman
Vice President, Forensics & Consulting

March 2010

THE IMPORTANCE AND CHALLENGE OF BACKUP TAPES FOR ELECTRONIC DISCOVERY

Before the 2006 amendments to the Federal Rules of Civil Procedure, and prior to landmark rulings such as *Zubulake v UBS Warburg*, paper was still king in the litigation discovery process. Electronic data was still mostly the domain of technologists in back office IT departments. Most lawyers were far less educated about litigation support technologies and judges were much more willing to rule that producing electronically stored information (ESI) was too burdensome.

Today ESI has taken center stage, and the legal framework for handling it races to keep pace with the ever-changing technology landscape. More seasoned and educated legal teams now routinely file comprehensive and complex discovery requests. Judges now tolerate few excuses for failure to produce ESI in accordance with court mandated deadlines, and are more likely to draw an adverse inference, apply sanctions or even resort to a summary judgment in situations of discovery misconduct.

It can be hard to keep up with advances in technology and the discovery issues raised by them. Today, lawyers must consider “metadata,” “social media,” and “cloud computing”¹ along with more traditional issues. This paper focuses on the unique discovery challenges posed by backup tapes and includes a comparison of findings from actual discovery projects. The case study demonstrates how new technology reduces traditional tape discovery costs by 75 percent and project time by 50 percent.

TAPES AS A SOURCE OF ESI

Backup tapes have long been a mainstay of electronic data storage, and thus a significant source of ESI. A single tape stores hundreds of gigabytes or even terabytes of data. The number of tapes in any given organization varies, but inventories of hundreds, thousands, or even tens of thousands of tapes are common. Not surprisingly, tapes are subject to discovery, particularly when they represent the sole source of requested records, or when other sources are deemed inaccessible². Consider that routine data destruction policies can eliminate other ESI sources. Failure to preserve data from backup tapes, intentional or not, could constitute spoliation in the eyes of a court or regulatory body³.

A BURDEN LIKE NO OTHER

E-discovery experts consider backup tapes to be one of the most challenging sources of ESI. Using traditional methods, restoring data from even a single tape can cost hundreds or thousands of dollars. For large tape collections, the cost can run into the millions. But cost is only part of the challenge. Technical challenges, which can be a major issue for timely preservation and production, are another problem.

Until recently, restoring backup tapes was onerous because doing so required recreating original backup environments. For Microsoft Exchange, the most widely used email system in corporate America, collecting data from backup tapes has traditionally required the original backup environment, including exact models, versions, and patches for all servers, operating systems, tape drives, and backup software. If the same environment is still in use, the restoration process is likely to disrupt daily business and IT operations. But more commonly, the

¹ Readers interested in e-discovery and cloud computing should consider viewing a recorded webinar in which I was a panelist on this topic: [The Challenge of Preserving and Collecting Evidence in a Cloud](#), Virtual LegalTech (Nov. 19, 2009), Panelists [Craig Ball](#), [Tom Morrissey](#), [Jeff Fehrman](#).

² Consider Judge Scheindlin's recent opinion in [Zubulake Revisited: Six Years Later](#), as posted on *JDSupra* by [Doug Cornelius](#).

³ Consider [Morgan Stanley to Pay \\$12.5 Million to Resolve FINRA Charges that it Failed to Provide Documents to Arbitration Claimants, Regulators](#), Financial Industry Regulatory Authority (FINRA), News Release (Sept. 27, 2007).

same environment is no longer in use, as systems have been replaced or upgraded, and thus must be recreated, a process often akin to scaling Mount Everest – it can be done but is a task of monumental proportions. Meanwhile, the court’s ESI production deadline looms in the background.

IT DOESN'T NEED TO BE THIS WAY

Backup tapes used to be every litigator’s nightmare, and today continue to keep many in the legal profession awake at night. But it doesn’t need to be this way. There are technology solutions available that greatly reduce the time and cost associated with e-discovery cases involving backup tapes.

USING TECHNOLOGY TO EASE THE DISCOVERY BURDEN

When a technology challenge is pervasive enough – as tape restoration can be – entrepreneurs step in to develop better solutions. As a trusted provider of e-discovery services, Integreon regularly reviews new products and services so that we can adopt useful ones. This allows us to offer our clients the latest in best practices and best-of-breed tools to ensure we deliver a streamlined, defensible, and cost-effective discovery process.

Recently, Integreon adopted Index Engines technology which enables the direct indexing and collection of data from backup tapes. A specific circumstance where Integreon employed this technology allowed us to compare the older tape restoration process with the newer process using Index Engines. In 2005, before Index Engines was available, we assisted a client on a matter involving many backup tapes. Then in 2009, after Integreon had adopted Index Engines, regulators reopened the case and the client needed to review the same tapes, plus additional ones. Below is a summary of our observations about the two projects, illustrating the importance of staying abreast of technology developments so that the right technology can be applied in each client situation.

A TALE OF TWO DISCOVERY PROJECTS

Project One in 2005

In 2005, our client received a regulatory request for responsive data from email sources stored on 480 backup tapes. These tapes were monthly full backups of corporate Exchange servers and other non-Exchange data.

Rather than recreating the backup environment, which would have been hugely burdensome, Integreon utilized Ontrack[®] PowerControls[™] to restore only the specific custodian mailboxes (user accounts) required from the collected tapes. A total of 56 mailboxes were identified and data covering a four year period was extracted.

PowerControls only processes Exchange databases⁴, so the tapes with non-Exchange data had to be cataloged manually to identify any non-Exchange email for the required custodians. This proved extremely tedious, requiring a substantial number of Integreon resources, and consequently added significant cost for the client.

The PowerControls’ restoration resulted in 4 terabytes (TB) of mailbox data. This is a huge amount of data, even by 2010 standards. To narrow the scope further, Integreon used other tools to index, de-dupe, and search the restored content, which reduced the data set down to 1.5 TB. The entire project took almost two years to complete and cost over \$8 million, including processing, hosting, and review.

⁴ Based on PowerControls reliance on Microsoft’s [MAPI \(Messaging Application Program Interface\)](#) technology.

PowerControls was the right tool at the time because it offered the ability to reach into backup tapes to identify and restore the specific custodian mailboxes, without having to recreate the original backup environment.

The following diagram illustrates the process used with PowerControls.

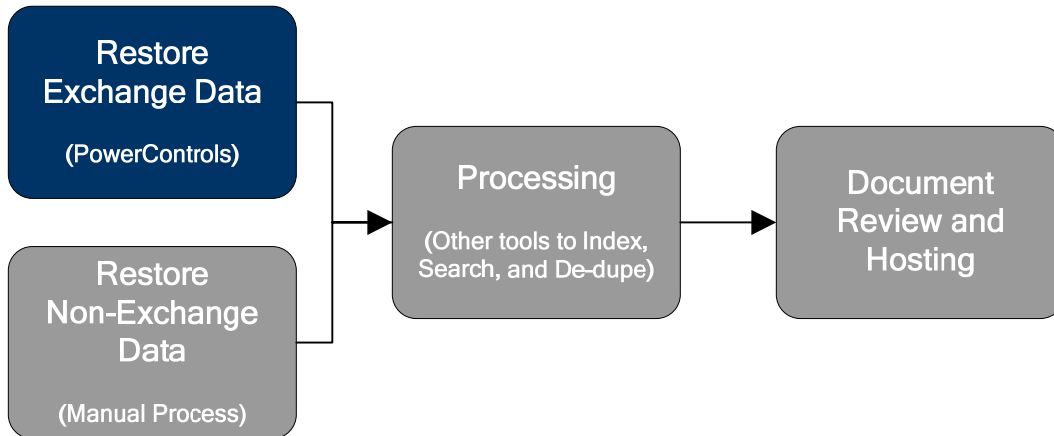


Diagram 1: Discovery Process for 2005 Project with PowerControls

Project Two in 2009

In 2009 regulators reopened the case, which required once again accessing the same backup tapes. Integreon advised the client on the latest available processes and tools, including Index Engines. The client selected Index Engines and decided to expand data collection to include PST, OST, and MSG file types and other Exchange content in order to develop a more comprehensive collection. The decision expanded the project from the original 480 monthly backup tapes Integreon first processed in 2005 to more than 800 tapes.

Using Index Engines technology, the Integreon team cataloged and indexed all of the tapes, which contained over 35 TBs of data, and made that content searchable. The process did not require the team to recreate the backup environment or even to access the data through the original software that created it. While it was still necessary to physically mount each tape on an Index Engines box, the effort and time to do this was minimal.

Beyond dramatically simplifying the restoration process, another key advantage of Index Engines is that the technology also automates the processing of non-Exchange backup data. This eliminated the need for the manual cataloging and processing as required in the first project in 2005.

Using Index Engines search capabilities and the criteria developed in conjunction with the client, the Integreon team searched metadata and content to eliminate many non-relevant emails. The resulting data set included only 500 gigabytes (GB) of relevant data from 800 tapes. This is in sharp contrast to the 2005 project, which resulted in 4 TB of data from 480 tapes. The time and cost savings were substantial as well, with the 2009 project costing just under \$2 million over a nine month period, as compared to the two year, \$8 million project in 2005.

The following diagram illustrates the process used with Index Engines.



Diagram 2: Discovery Process for 2009 Project with Index Engines

Why the Differences?

Comparing the two projects from 2005 and 2009, there are three important differences which are all a result of the capabilities of the Index Engines technology.

1. Processes that had to be performed manually in 2005 were automated in 2009, which accelerated the project and reduced costs.
2. The processing sequence was reversed to minimize the effort expended on irrelevant data. In 2005 the team had to first restore the data and then process it further in order to narrow the scope. In 2009 using Index Engines, the team was able to first narrow the scope (even more precisely than in 2005) and then extract only the documents required.
3. More tapes could be processed in a shorter period of time, enabling more comprehensive results. Index Engines also does not rely on MAPI-based technology, which further enhanced the results⁵.

CONCLUSIONS

Applying the right technology in each discovery situation is critical to controlling costs and achieving the best outcome. While backup tapes will continue to be a challenging and costly medium, solutions like Index Engines can result in savings that could mirror the 75 percent reduction in cost and 50 percent reduction in elapsed time that the Integreon team demonstrated in the 2009 case.

The following table summarizes the results that the Integreon team demonstrated in our 2005 and 2009 cases.

Projects Summary & Comparison		
Project One – 2005	Project Two – 2009	2009 vs. 2005
480 tapes/56 custodians/1.5 TB from 4 TB (Including Processing, Review and Hosting)	800 tapes/500 GB from over 35 TB (Including Processing, Review and Hosting)	320 more tapes 31 more TB processed
Cost: \$8 million	Cost: \$1.9 million	25% of the cost
Time: 18 months	Time: 9 months	50% of the time

⁵ MAPI technology cannot access email with any broken message links, such as email scanned and corrupted by virus software, or email that has been deleted into a mail server's "dumpster".

FINAL THOUGHTS

Integreon routinely monitors e-discovery technology developments. We quickly saw the value of Index Engines and partnered with them in 2008. We have since used the technology extensively. Available as our [Advanced Tape Discovery™](#) service, we have found that Index Engines is radically different from other tools, such as PowerControls, and we have seen tremendous savings in both the time and cost of processing backup tapes.

Leveraging technology to solve complex discovery challenges, such as those associated with backup tapes, is an important part of Integreon's e-discovery offering. Our consultative approach to selecting, implementing, and managing discovery solutions enables our clients to choose the right services for each project. To learn more about our Advanced Tape Discovery service or our entire range of integrated discovery solutions, please visit us online at www.integreon.com/discovery.

ABOUT THE AUTHOR

Jeffery C. Fehrman is Integreon's Vice President of Forensics & Consulting, based in Arlington, Virginia. With more than 13 years in the electronic evidence and information technology fields, Mr. Fehrman specializes in providing consulting and investigative services to law firms, corporations, and government clients. Fehrman has lead numerous forensic investigations in theft of IP, web defacement, Freedom of Information Act (FOIA) requests, and many other areas. He is a subject matter expert on electronic evidence and frequently speaks on topics such as e-discovery innovations, and the e-discovery obstacles facing corporations and law firms. Fehrman has also developed accredited courses approved by the Continuing Legal Education (CLE) boards in the States of Illinois, Georgia, New York, Pennsylvania, and Virginia, and has trained hundreds of attorneys in the technical aspects of computer forensics and electronic discovery.

ABOUT INTEGREON

Integreon is the global e-discovery and Legal Process Outsourcing (LPO) partner trusted by the world's leading corporations. It is the only legal services provider in the world to receive top marks in both the Socha-Gelbmann Electronic Discovery Survey and the Black Book of Outsourcing LPO Survey. To most efficiently, accurately and defensibly identify responsive data from even the largest data sets, the company provides an integrated approach to all stages of the discovery process, across scalable operations and technology: unified project management, consulting, forensics, data analytics, processing, hosting, and review services. Experienced document reviewers, working on-site at client offices or at any of Integreon's secure review centers worldwide, following ISO 9001-certified best practices, deliver high-quality, repeatable review results. For more information about Integreon, please visit www.integreon.com.

ABOUT INDEX ENGINES

The patent-pending Index Engines discovery platform is the only solution on the market to offer a complete view of electronic data assets. Online data is indexed in-stream at wire speed in native enterprise storage protocols, enabling high-speed, efficient indexing of proprietary backup and transfer formats. Index Engines' unique approach to offline records scans backup tapes, indexes the contents and extracts relevant data, eliminating the time-consuming restoration process. Index Engines provides the only comprehensive discovery platform across both online and offline data, saving time and money when managing enterprise information. For more information on Index Engines, please visit www.indexengines.com.