

White Paper

Methodology for Historical Records Management

By Jim McGann, Index Engines
October, 2008

Records and Information Management (RIM) is Only as Strong as the Data

Information is the life blood of a company. Valuable information is generated every day, data that will provide your organization with a competitive edge. Capturing, storing and organizing this information into a records management platform or archive has become increasingly common, allowing it to be leveraged in order to make more informed decisions. Those companies that have not yet implemented a records management strategy are now jumping on board due to more compelling issues; corporate compliance and risk of litigation. Legal and compliance issues have driven the need for a comprehensive records management strategy since federal and state requirements have become more stringent with respect to corporate data. Federal Rules of Civil Procedure (FRCP), Sarbanes-Oxley, and other recent regulations mandate how electronically stored information (ESI) such as contracts, spreadsheets and email are handled internally, and more importantly produced to the courts when required.

Those who do not have a records management or archive strategy in place today are quickly making plans for one tomorrow. The repercussions of not having your information assets managed properly are steep fines and possibly legal ramifications. So where do you begin? The obvious place is with current ESI. Data being produced today, or that was created over the past few months, is typically online within your organization's network and is easily accessible. Email archives, records and content management solutions, as well as litigation hold platforms all are designed to pull data from your online network and consolidate it for safe keeping. But what about older records? Data that is no longer online presents a complex challenge. How do you account for historical ESI, and ensure that these older records are included in your strategy?

Making the Case for Historical Records

It can take years for a lawsuit to reach the courts. The legal process is slow and cautious. Many conversations and negotiations occur before evidence must be produced. When the time comes to dig through mountains of ESI, locating the relevant data could be very arduous. A case in the courts today, is typically about a situation that occurred years ago. Imagine putting all your team's effort into building a world class platform only to find that the desired data was created before it was launched and was not included in the online repository. Most records management and archive solutions are only useful for data generated from the time they are launched. Historical data is typically not included as it is archived off the network and very difficult and expensive to access. Therefore historical ESI is neglected. However, this historical data contains valuable information - information that is required by your legal team in order to support critical litigation.

Strengthen the RIM Process by Incorporating Historical Records

When considering a records management or archive platform it is important to not only include current online information assets, but also historical ESI that is typically archived off the network. Importing both current and historical data will provide the comprehensive view of ESI required by your legal team in order to intelligently defend against litigation. Knowing what happened in your business yesterday, last month or even last year is important. However, knowing what happened a few years back can make or break a case.

So how do you access this historical data? Where is it? How can it be ingested into your records management and archive platforms? Historical information is typically comprised of ESI that has been backed up by your data protection team and archived on backup tapes. These tapes are typically stored offsite for safekeeping and serve a purpose for only a very short time. Their only purpose is to archive data for a specific retention period, typically less than 30 days, in the case of a catastrophic failure of the data network. In the case of such failure, data can then be retrieved and no critical business information is lost. Once this retention period is over the backup tapes are typically left in offsite storage, having lived out their purpose, and newer tapes arrive to take their place alongside the older versions. Tapes pile up with no specific intent in mind - until now.

Thousands of historical backup tapes are sitting in offsite storage collecting dust. Imagine if you could backfill your records management system with the relevant data from these tapes. This would allow a complete picture of your enterprise data assets - a view that goes back in time and is easily accessible by your legal team to support ongoing litigation. This data would include email from old employees that are no longer with the firm, valuable project research data, even contracts and agreements that were the basis for existing product lines. Capturing this data, that is typically a burden to collect during a court case, and seamlessly integrating it into your repositories will provide the knowledge that will satisfy even the most demanding legal team.

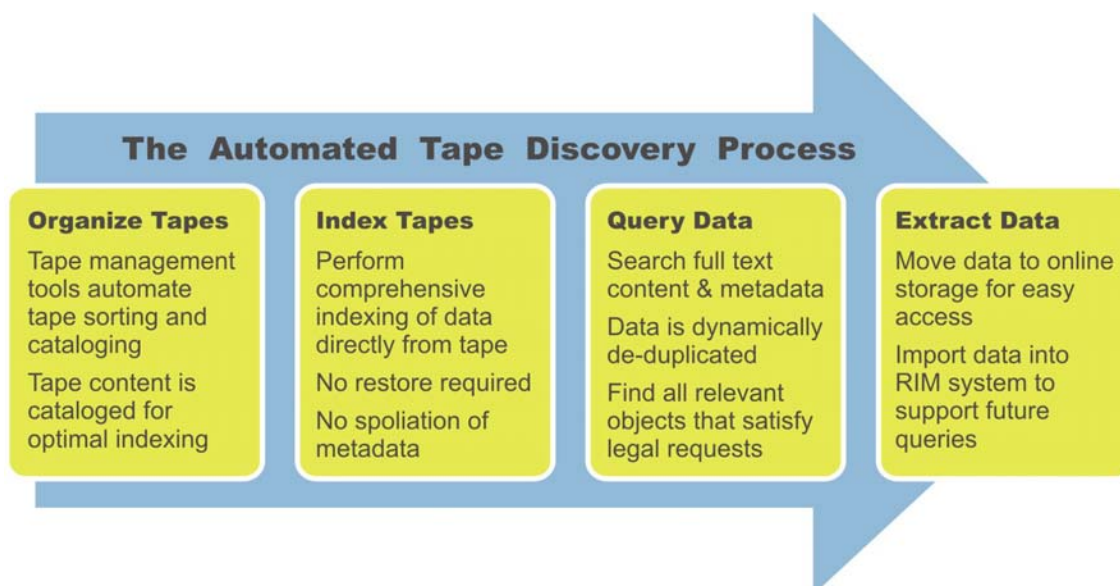
Why is this historical data so difficult to access? When IT backs up data onto tape they use software that optimizes the process, making it efficient and timely to get large volumes of ESI off the network and onto tape before users come back to work. In doing so, data is placed in a proprietary container specific to the backup software in use at the time. Access to data in this container, or backup format, is easy if the backup environment is still in place. However, as the environment changes, this data becomes locked away. Additionally as the volume of tapes increase, it becomes more and more difficult to know what is on these tapes. A complete tape labeling system would be helpful, however this is typically neglected and tape content quickly turns into a mystery. Thus when a tape is retrieved from storage and the IT department, or outside tape restoration partner is tasked to find specific content, you soon learn it is not a simple process.

Apply RIM Policies to Tape Data

How do you get access to data on tapes? How can you backfill your archive or records management platform with the historical ESI locked away on offline archives? And how can you do this without breaking the bank? First off, you only need the relevant data. Backup tapes consist of large volumes of duplicate data. It is estimated that over 50% of data on tape is redundant. Additionally you can remove irrelevant data, such as email exchanges for the purpose of meeting for lunch as well as spam email. Typically less than 10% of data on historical backup tapes is data appropriate to include in a content management platform. Let's investigate how you get access to this relevant data.

Accessing data on backup tapes has traditionally required that the entire contents of the tape first be restored. Restoring data from tape requires that the environment used to create the tape be recreated, even if it is obsolete, and then the data can be brought back online. This process is expensive and time consuming. In fact, it can require forensic experts with specialized knowledge of these proprietary backup environments. This laborious process is precisely the reason why in the past records management and archives have not been backfilled with relevant data from offline tape.

New technology has changed the way data on backup tapes is accessed. Technology now exists that can read these proprietary backup formats, understand what ESI is stored on tape, and then pull relevant content into your archives and records management platforms quickly and economically, without having to recreate the original environment. The relevant data is extracted in its original format, with all the metadata intact, without ever having to restore the tape.



The Stages of the Automated Process for Collecting Historical ESI for RIM Systems

Index Engines has cracked tape formats, allowing rapid access to files and email locked away on tape. You do not need to be an IT specialist, understand backup formats, or know anything about data forensics to implement this technology. A simple spin of the tape will tell you everything you need to know. For example; you are looking for all email exchanged between the executive team, or documents related to a specific project. You can simply scan a pile of tapes and an index of the metadata and content is quickly generated. Searching this index will provide a unique list of all ESI that is responsive, all duplicate files and email ignored. No longer do you need to restore all contents from tape to commence the discovery process. Now you can scan tapes and investigate the contents without ever restoring anything. Once all the responsive data is found, you can then extract this data from tape and load it into your archive or records management platform quickly and painlessly. No longer is data contained on tape a burden to extract. Index Engines is allowing split second access to the historical records needed for a comprehensive view into *all* your ESI.

Get Litigation Ready®

Backfilling your records management or archive platform creates a thorough repository of all your organization's electronically stored data – not just the information from today on. Historical records will help provide a more comprehensive insight into the digital evidence required by your legal team. This complete view will allow them to more intelligently argue cases and defend litigation now and in the future. Without access to data stored on tape legal team's ability to make fully informed decisions is hampered. If you are implementing a records management strategy that does not include historical data contained on backup tapes your system will be incomplete. Your organization will suffer from both the liability of idle data on tape and the absence of access to stored historical knowledge. A sound approach to records management includes data from both yesterday and today, leaving no gaps in your repository. Unlock the knowledge contained on backup tapes and ensure your organization is leveraging the full power of your electronic records.