

# White Paper

---

## **How Archiving Reduces the Cost and Complexity of “Reactive” eDiscovery**

*By Brian Babineau*

**April 2010**

---

This ESG White Paper was commissioned by Symantec and is distributed under license from ESG.

## Contents

Introduction .....	3
The Intersection of Electronic Discovery and Archiving .....	3
Beyond Backup Tapes .....	3
The Evolution of Archiving Solutions .....	4
The Transition to In-sourcing .....	5
Centralization Drives Efficiency .....	5
Before the Request Arrives .....	5
Kill Three Birds with One Stone .....	6
Mitigate Risk .....	6
Dealing with Data Outside the Archive .....	7
Targeted Collections .....	7
Minimize Software and Consulting Expenses .....	8
Repository of Record .....	8
Controlling Distributed Information .....	8
Key Considerations .....	9
The Bottom Line .....	9

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

After receiving a discovery request, IT and legal departments must identify, preserve, and collect potentially relevant content (There is precedent finding that companies should commence the preservation process if they can reasonably assume a discovery request is coming.).<sup>1</sup> When a significant proportion of potentially relevant electronically stored information (ESI) resides in a central, searchable location where it can be identified, collected, and protected, eDiscovery is much easier. This is one of the fundamental benefits of a purpose-built archive designed specifically to retain information for compliance, legal, or business reference reasons. What most companies do not realize is that an archive also delivers measurable value in situations where the scope of an electronic discovery request involves data that is distributed across the organization.

When IT and legal departments create an archive and know exactly what it contains, they can use the time they would have spent searching for that data to execute more efficient, targeted identification and collection processes of non-archived ESI wherever it may exist. As an example, if a company archives all e-mails, it would be redundant to collect data from the primary messaging environment. An archive can also serve as a repository of record for legal matters—by moving data into an archive after it is collected, users can then take advantage of centralization benefits such as the ability to implement legal hold policies, monitor and report on any information access activity to support chain of custody requirements, and complete “early case assessments.”

Archiving is an extremely versatile information management solution that can aid in several different business processes ranging from storage management, to electronic records management, to electronic discovery. Though more efficient, streamlined electronic discovery is just one benefit of archiving—the others include compliance, governance, and IT resource optimization—it is extremely important and still often misunderstood. This paper attempts to identify and clarify the benefits an archive can provide in the initial phases of the electronic discovery process and look at some of the features companies should look for to ensure those benefits are maximized.

## The Intersection of Electronic Discovery and Archiving

### Beyond Backup Tapes

To completely understand the value of an archive in electronic discovery, it helps to understand why companies started using archiving to support legal processes in the first place. Entering this decade, regulators and litigators targeted IT systems as sources of evidence as businesses transitioned into the digital age. The easiest place to find historical information was usually on backup tapes, as this was the most central storage point in IT. The problem was finding relevant data across hundreds or thousands of backup tapes in a relatively short timeframe. The process proved to be extremely difficult and expensive. Even if a company could afford to restore data from backup tapes, the complex operations barely left time for attorneys to prepare for Meet & Confer sessions or develop effective case strategies.

In addition to the restore times, IT and legal had to remove any tapes containing potentially relevant information from traditional backup rotations in order to address preservation requirements. As the amount of matters involving ESI increased, tapes on legal hold began piling up—if any of these were lost or overwritten, a company could be accused of hiding or maliciously deleting evidence, leading to an adverse opinion or increasing the likelihood of an unfavorable outcome. There are several notable cases of this, including *Perleman v. Morgan Stanley*, where companies have been forced to pay hundreds of millions of dollars in compensatory and punitive damages because they failed to identify all information on tape or properly preserve it, or both.<sup>2</sup>

The increasing cost and risk of electronic discovery drove the need to make corporate archives more accessible, forcing companies to move away from traditional data retention implementations (i.e., saving backup tapes for longer periods of time). More companies began to realize that archiving differs from backup and involves copying or moving data from a production application to a separate environment for compliance, governance, electronic

---

<sup>1</sup> Gibson, Dunn, & Crutcher, 2009 Mid-Year Update on E-Discovery Cases, July 2009.

<sup>2</sup> <http://www.infoworld.com/d/developer-world/ripple-effect-court-cases-means-new-rules-it-947>.

discovery, and business reference. A separate environment is critical as it enables organizations to manage information—including indexing the content as well as assigning and enforcing retention policies—without impacting production applications. Additionally, users can save archived data on less expensive storage infrastructure, keeping more information online and, most importantly, accessible at a lower price point.

## The Evolution of Archiving Solutions

Purpose-built archiving solutions facilitate the movement or copying of data from a production environment to a separate one. Consistently executing these operations is critical if attorneys are to rely on the archive as a complete “system of record.”

In addition to keeping more information accessible, several other purpose-built archive solution capabilities improve electronic discovery processes including:

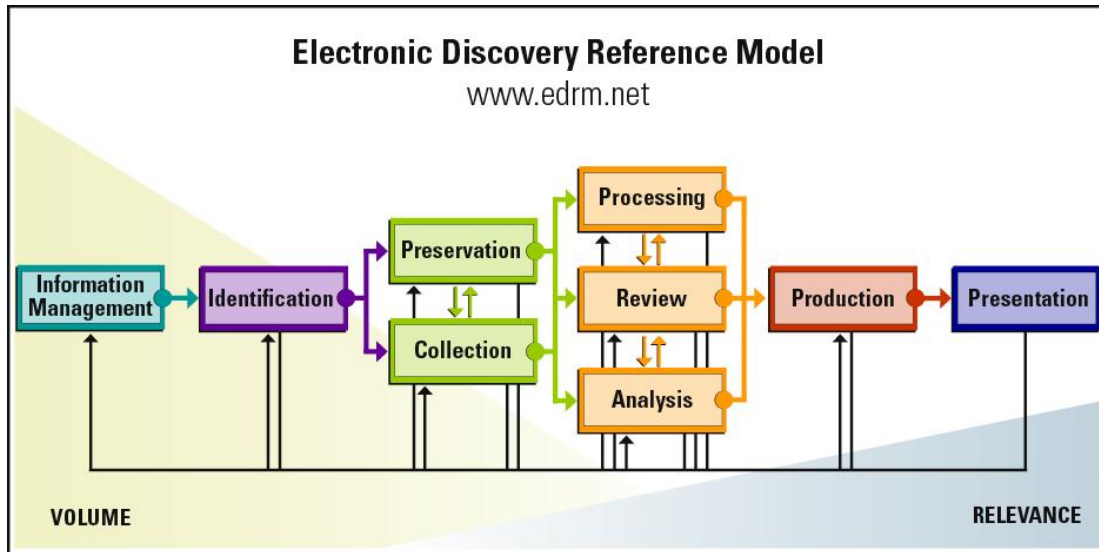
- **Deduplication.** As archive solutions ingest messages from the primary application environment, they also check content that has already been stored for duplicates. Rather than saving a file, message, or other data type multiple times, the solution creates a reference pointer between the redundant data and its match in the archive, saving the data only once. As a result, the archive is smaller, which facilitates faster searches when a discovery request arrives.
- **Indexing.** All content within the archive is indexed, enabling IT and legal to execute queries focusing on specific custodians, date ranges, or keywords. By having information readily searchable, attorneys and IT can commence eDiscovery activities immediately as opposed to waiting for the indexing process to occur—a task that can take several days depending on how much content there is and where it is located. In turn, attorneys can begin reviewing and assessing information much more quickly.
- **Retention management.** Companies can establish a retention policy determining what data needs to be kept and for how long. Simply identifying the data that needs to be saved reduces the amount of information that has to be searched after a receiving a discovery request. If messages are to be saved, policies determine when the e-mails should be moved to the archive and how long the data will be kept there. This enables companies to consistently delete data when a retention policy has expired, minimizing the amount of information that lingers and remains discoverable.
- **Tagging.** When data within an archive needs to be placed on legal hold, relevant messages can be tagged in a number of ways to do things like prevent the content from being deleted or modified until a matter is resolved. The actual execution of the legal hold is important as well. Some archive solutions copy content from the primary archive repository to a different one, increasing the amount of data that exists and must be managed. Also, there are now two copies of the same content, making it hard to properly delete the data when it comes off legal hold. When a “tag” is added to the data, no copies are necessary – the messages are managed with a legal hold “in place.” The “in place” option becomes much more manageable when companies are dealing with several different, overlapping legal hold requirements on the same data—some of which can last for several years.

Some purpose-built archive solutions have add-on electronic discovery applications that enable users to add tag relevant information during the search process and construct workflows associated with those tags. Tagging also enables companies to denote data as non-responsive, allowing them to focus on managing and organizing responsive information while the workflow facilitates initial analysis and review tasks for a particular matter. Tagging can substantially reduce the initial review process as attorneys do not have to manually read through all the content marking each item along the way. The time savings reduces costs as the initial review process is typically handled by outside counsel (or contractors hired by outside counsel) with their own billing cycles.

## The Transition to In-sourcing

Electronic discovery is a complex process with several related steps (see Figure 1). When only a few key corporate data sources and matters involving ESI were the exception rather than norm, companies could outsource the process to specialized consultants, legal service providers, and external counsel. Now, outsourcing is much more expensive due to the growing number of cases involving electronically stored information and increasing data volumes—consultants and attorneys charge by the hour and legal service providers typically price by the amount of data they manage per matter.

Figure 1. Visual Representation of the Electronic Discovery Process



Source: edrm.net.

It is unlikely that companies are going to eliminate outsourcing altogether as there are some steps within electronic discovery—including final review and production—that require highly skilled experts and specialized technologies. There is, however, a trend toward organizations bringing certain steps of the electronic discovery process in-house by investing in technology such as purpose-built archiving solutions. These solutions automate many of the preliminary steps in the electronic discovery process, enabling tasks to be completed by internal resources. An October 2009 report released by Fulbright & Jaworski, a well-recognized global law firm, confirmed this trend as approximately half of corporate respondents plan on in-sourcing some aspects of electronic discovery process in the near future.<sup>3</sup>

## Centralization Drives Efficiency

### Before the Request Arrives

Just by having a purpose-built archive solution in place, an organization is already improving electronic discovery processes even if they have yet to receive an inquiry. As mentioned, all content is indexed, allowing it to be searched as needed. These searches can be executed against a smaller corpus of data if companies choose to delete information when retention policies expire. This is possible as archiving facilitates consistent data expiration (deletion) to actually reduce the amount of ESI that could potentially be discovered—a task that is allowable under the amended U.S. Federal Rules of Civil Procedure so long as the data is not currently under legal hold or no longer needed for other compliance purposes.

Far too many organizations believe that archiving involves the proper retention of information. This is a limited perception as companies can use archive solutions to determine what to keep via business rules that typically

<sup>3</sup> Fulbright & Jaworski, *6th Annual Litigation Trends Survey Report*, October 2009.

cannot be executed (or are too disruptive to execute) on primary systems. Further, archiving also facilitates proper data expiration when it is no longer needed for legal, compliance, or business reference purposes. And, more importantly, when data is deleted after the retention policy expires, companies do not have to worry about other copies of the same message or file lingering within the archive because the content is deduplicated as it enters the archive.

### **Kill Three Birds with One Stone**

Upon receiving an electronic discovery request, most attorneys immediately call their IT departments. After identifying where data may be, attorneys notify certain custodians or data owners that relevant information needs to be placed on “legal hold.” In many cases, individual custodians may preserve data themselves on their PCs until in-house counsel deems it needs to be collected. IT then usually copies the information from any relevant sources, including employee PCs, corporate file shares, primary applications, and backup tapes. During collection, IT usually gathers more information than is necessary. Suppose a complaint involves an employee dismissal – the worker’s attorney may request certain information created within a certain date range from the individual’s supervisor. IT “images” (take a full system copy) the supervisor’s PC and copies all data from a file share that the supervisor had access to, even if some of the data may not be relevant to the request. When collection is complete, IT usually makes another copy of the data and sends it to an external party, such as external counsel or a legal service provider, further ensuring preservation.

Archiving enables IT and legal to execute identification, collection, and preservation nearly simultaneously. Most solutions have “super user” access roles allowing authorized individuals to search the entire archive, expediting identification. Depending on the matter as well as a solution’s capabilities, users can tag, copy, or export search results, facilitating collection. And, any content collected can be placed on legal hold within the archive to satisfy preservation requirements.

All of these electronic discovery steps can be executed against existing data and any new ESI that enters the archive. Attorneys can save searches and schedule them to run periodically to see what new data meets a specific discovery request. Some purpose-built archive solutions have automatic classification capabilities that tag any new data ingested into the archive triggering the designated search criteria. In continuing with the example above, if a company places a supervisor’s e-mail on legal hold, any messages entering the archive sent from or to the supervisor can be automatically tagged with a case name and set with an indefinite retention policy.

### **Mitigate Risk**

By compressing the initial phases of the electronic discovery process, archiving gives attorneys earlier insight into case facts and timelines. This enables attorneys to prepare for Meet & Confer sessions and initial arguments, and to strategize their approach to a specific matter. Often referred to as “early case assessment,” the ability to understand more about a matter sooner can also guide attorneys in making “settle” versus “argue” decisions, potentially avoiding legal expenses and costly unfavorable outcomes.

Those purpose-built archive applications that have an add-on electronic discovery application can also facilitate initial reviews to further bolster insight into case facts earlier. Using customizable tags with associated workflows, attorneys can designate certain content as responsive, unresponsive, privileged, or another category of their choosing. Doing so prevents the unnecessary export of irrelevant data to service providers for final processing, potentially saving a company millions of dollars. Additionally, attorneys can filter on any tags before exporting the information—either to a service provider or opposing counsel—reducing the chance of accidentally sharing privileged data.

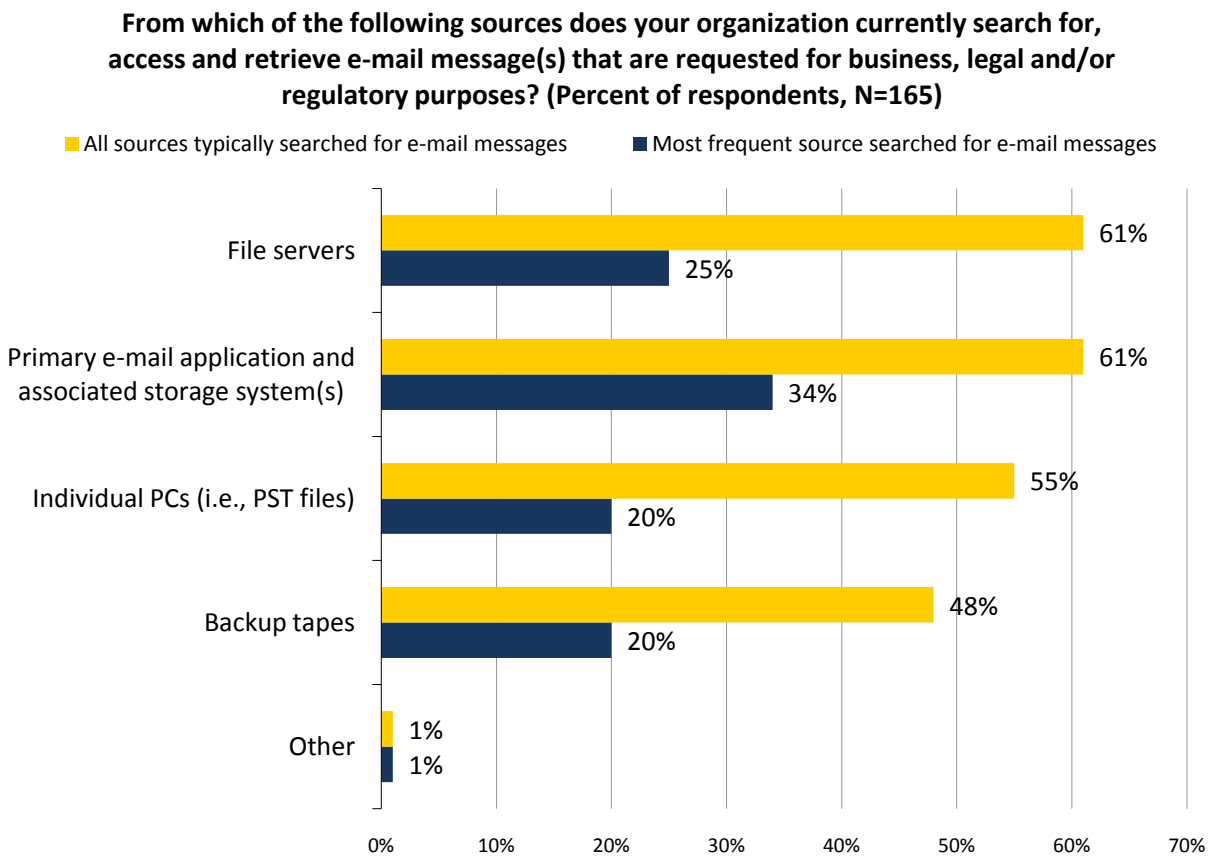
When an archive has an add-on electronic discovery application, attorneys can become self-sufficient. They can execute searches and tag information using the application interface without having to involve IT. This flexibility is useful if the initial phases of discovery involve sensitive case subject matter. It also is useful if attorneys are using the application to support an internal investigation where they want to limit the amount of individuals involved in the process.

## Dealing with Data Outside the Archive

### Targeted Collections

The benefits pertaining to electronic discovery when the data is in the archive are relatively straightforward. However, archiving also helps IT and legal manage discovery of data that may be outside of a central repository. First, companies can determine what is in an archive and what isn't, reducing “guesswork” collections and the cost and risk associated with over collection. Continuing with the supervisor above, if a company already has e-mail archiving in place, then they can choose not to collect any PSTs from a PC or any file share that contains message files (Figure 2 shows many of the places from which organizations currently collect e-mails because they simply are not sure where all the data is). The same holds true of other content types including file shares and SharePoint data that can be easily archived. Take, for example, a company that has a “financial planning” file share site where data is archived after 90 days. If a discovery request involves data created with the last month, IT and legal can focus their efforts on the file share itself. In contrast, if the inquiry specifies data created two years ago and archive retention policies call for three years, an organization can commence collection within the archive without having to revert to backup tapes.

Figure 2. Sources Most Frequently Searched for Required E-mail Messages



Source: ESG E-mail Archiving Survey – Preliminary Results, March 2010

One of the more significant time savings resulting from targeted collections is the imaging of PCs. As mentioned, collections often start with IT creating full copies of a named custodian’s PC—a process that can last hours depending on the size of the hard drive. What’s worse is that most of the data within a PC image, including application and configuration files, are not relevant. They simply increase the amount of data that is sent to an external service provider for processing and review. This seemingly benign task can be very disruptive, especially when remote office employees are involved. Attorneys and IT resources often have to be dispatched to various locations as network-based imaging is rarely possible given the amount of data.

## Minimize Software and Consulting Expenses

Due to the amount of corporate data that could be discoverable, many companies have turned to replication or copy tools to automate identification and collection processes when data is decentralized. Most of these tools, including those designed for PCs and file shares, are often priced by the amount of data analyzed rather than by what is actually “collected.” ESG spoke with a litigation support manager working for an 8000-person energy company who licenses upwards of five collection tools to address electronic discovery requirements—not an uncommon finding for a company that experiences a reasonable amount of inquiries in a given year. If a company can reduce the amount of data that sits outside of the archive, they can cut license fees associated with these collection tools while getting all of the benefits (centralized legal hold enforcement, reduction in service provider fees, quicker initial reviews and early case assessments, etc.) delivered by the archive.

For some companies, identification and collection efforts are so disruptive to normal business processes—in-house attorneys and IT have their day jobs to do—that they turn to external counsel or specialized consultants. For example, a company managing several concurrent electronic discovery processes may have e-mail data stored within a primary e-mail application’s desktops and file shares (due to personal archives) as well backup tapes. If IT cannot deal with all of the requests or easily access the data—a problem that often occurs when backup tapes are involved—the company is likely to use consultants to project manage and execute all of the collections. Companies can minimize reliance on these external experts and the burden on internal IT departments by centralizing data in the archive.

## Repository of Record

Even if a company chooses to archive only one content type such as e-mail or SharePoint sites, they can use the archive as a matter repository where data is preserved. After distributed data is identified, IT can copy it into an archive where retention policies can be extended until a matter is resolved rather than shipping the data offsite. Being proactive with preservation efforts may actually pay dividends in the long run according to a recent report issued by Gibson, Dunn, & Crutcher, a leading global firm with an electronic discovery practice.

*One interesting theme emerging from the duty-to-preserve opinions is the use of e-discovery offensively. Typically, a party establishes a data preservation protocol in order to defend or shield itself from accusations of discovery shortcomings and potential sanctions resulting from them. However, a party that is conversant in e-discovery can also use it to assail an opposing party if it fails to meet its obligations.<sup>4</sup>*

At a time where companies may be expected to put data on legal hold even before receiving an inquiry, having a place to centralize and efficiently manage such a process can be extremely useful.

## Controlling Distributed Information

When companies implement an archive and move or copy data to it, it actually mitigates the amount of data distributed across the organization. As an example, an organization may choose to archive e-mails, removing the need for employees to create personal archives (PST or NSF files) and save them on PCs or file shares. The same holds true for SharePoint sites—by storing fresh data within SharePoint itself and archiving older information, application performance remains stable and backup processes are simplified, removing the need for users or IT to create new sites because existing ones get crowded. In short, archiving can help an organization gain control over its information.

---

<sup>4</sup> Gibson, Dunn, & Crutcher LLP, *2009 Mid-Year Update on E-Discovery Cases*, July 2008.

## Key Considerations

Organizations need to understand the feasibility of an archive—not every piece of corporate data can or needs to be put into an archive for electronic discovery purposes. However, most companies can afford to expand how they are using current archives today by adding file system, SharePoint, and even database information into current e-mail archive repositories. From that perspective, choosing an archive solution that scales and supports all of these content types is critical.

When companies are centralizing information in an archive to facilitate electronic discovery, they will likely actively evaluate solutions that have a purpose-built electronic discovery application that enables the aforementioned tagging, centralized in-place legal hold, and first pass review capabilities. Companies may also want to consider an archive solution provider’s partnerships and technology integration with legal service providers. Such agreements allow for data to be exported from an archive directly into a service provider’s processing engine so it can be prepared for final review and production. The benefit to customers is a much simpler, more manageable chain of custody process.

Lastly, knowing that some data will always be outside the archive and discoverable, organizations would be wise to work with archive solution providers that can, at a minimum, help identify and collect this information. In some situations, a collection of offering may be separate from the actual archiving solution, but it does minimize the number of vendors a company needs to work with in order to complete ESI collections. Customers should keep in mind that these collections will be more targeted and not free-for-alls if they are archiving regularly. With both solutions coming from a single provider, collected data can be typically ingested into the archive to be indexed and made fully searchable for the current matter and any future events that may occur. In addition, legal hold can be applied to the data to secure it from inadvertent or intentional destruction.

## The Bottom Line

Archiving is not a new process to most IT departments, but new business requirements—particularly electronic discovery—have changed the way it needs to be executed. In fact, electronic discovery has become so commonplace that companies have no other choice but to implement purpose-built archive solutions to alleviate the burden on their legal and IT departments.

It is fairly straightforward to see how companies can expedite such a complex process—one that continues to become more complicated as the amount of data and matters increase—by automating identification, collection, and preservation tasks. What many don’t realize is that archiving also enables them to control data dispersion, manage other legal-related technology expenses, and target collection activities when data is not stored in the archive itself.



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