



WHITE PAPER

Implementing a More Efficient Information Archiving and Retention Strategy

By Brian Babineau

May, 2009

Table of Contents

Table of Contents	i
Introduction	1
The Need to Retain (the Right) Information	1
Electronic Discovery.....	1
Compliance	2
Business Intelligence	3
The Common Requirement.....	3
Old Methods Won't Work	4
Backup versus Archiving.....	4
Manual Archiving.....	4
Forced Behavior.....	4
The Common Pitfall.....	5
A Better Way	5
Purpose-Built Archive Software	5
What to Look For?.....	6
The Benefits	7
Conclusion	10

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. This ESG White Paper was developed with the assistance and funding of C2C.

Introduction

By now, many organizations should realize that archiving—the process of saving files, e-mails, and other content in a repository separate from the production environment—delivers several benefits including lower storage costs, improved primary application performance, and the ability to facilitate compliance with record retention regulations and automate electronic discovery (or electronic disclosure) operations. However, legacy archive implementations have revealed two major problems: First, companies do not have the proper mechanisms to archive the right information (business records, data relevant to an electronic discovery request, etc.) forcing companies to save more than they have to. This leads to the second issue: archive environments are rapidly growing out of control as customers do not have the right tools to manage the information once it is saved.

Companies need the ability to archive a subset of information and consistently enforce retention periods based on predefined policies. As archives get bigger, customers would also benefit from more precise archived information management. Automation of such tasks will allow organizations to extract greater efficiency from their archive implementations and facilitate the completion of eDiscovery processes. Automated archive and retention management can also significantly reduce storage expenses as companies save only what is needed, expire data when retention periods or legal hold requirements have been met, and move archived information to less expensive systems over the time.

Building an efficient information archive and retention strategy can be reduced to a three step process. First, an organization must realize that it cannot avoid retaining information in today's business climate. Record retention regulations, electronic discovery mandates, and the ability to improve decision making and productivity by making more data available to knowledge workers will force any organization, if they do not already, to save more data for longer periods of time. Second, a company must accept that existing information archive retention processes and supporting technologies are not the most cost effective means to balance storage costs with data accessibility. Specifically, using standalone backup software and tapes is a means to save data inexpensively, but the information is not readily accessible and therefore cannot be easily leveraged. Lastly, customers should become more familiar with purpose-built archive solutions that can provide a better means of archiving and retention. Some of these solutions have unique capabilities that automate the archive process by identifying which data to keep and, more importantly, what does not need to be archived. Further, companies should evaluate solutions that offer certain features such as the ability to set very discrete policies to manage the information during the retention period, including the storage system on which archived data is saved and when to expire the content.

This paper explains, in detail, these three steps so organizations can build more efficient information archive and retention management strategies.

The Need to Retain (the Right) Information

Electronic Discovery

In August, 2008, The American College of Trial Lawyers released a research report focusing on the issues associated with the discovery process as it pertains to the advancement of the American legal system. Nearly 1500 trial lawyers responded to a series of questions regarding discovery issues—some of which centered on electronically stored information (ESI). Approximately 60% of the respondents reported having cases that raise electronic discovery issues. Of that group, over 86% have issued or received a discovery request for ESI since the new U.S. Federal Rules of Civil Procedure (FRCP) went into effect in December, 2006.¹ It is also important to note that the United Kingdom also updated its Civil Procedure Rules in roughly the same timeframe, with much

¹ American College of Trial Lawyers, *Interim Report on the Joint Project of The American College of Trial Lawyers Task Force on Discovery and the Institute for the Advancement of the American Legal System*, August 1, 2008.

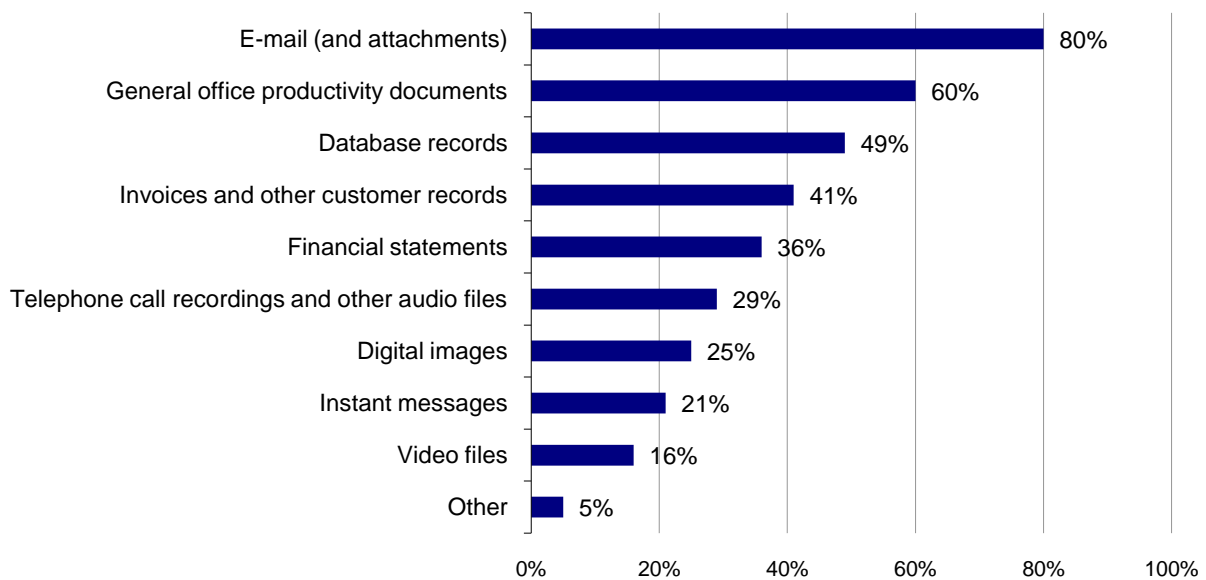
Implementing a More Efficient Information Archiving and Retention Strategy

of the changes centered on the disclosure of electronic information. Additionally, many organizations execute internal investigations and audits that follow the electronic discovery (or disclosure) process.

During the initial phases of the electronic discovery process, organizations must collect information from a variety of sources, including e-mail and file servers (see Figure 1). This data must be protected from alteration or deletion until it is deemed irrelevant to a specific matter or until a case is resolved. With legal and regulatory matters potentially lasting years, companies are often saving electronic evidence for long periods of time. And, once data is collected and preserved, corporate counsel must sift through piles of information, seeking a small portion of data that is relevant or 'responsive' to a discovery request.

FIGURE 1. RECORD TYPES REQUESTED DURING ELECTRONIC DISCOVERY EVENTS

To the best of your knowledge, which of the following record types has your organization been asked to produce in a legal proceeding or regulatory inquiry? (Percent of respondents, N=107, multiple responses accepted)



Source: ESG Research Report, 2007 E-mail Archiving Survey, November 2007

Compliance

Large Wall Street banks are not the only firms required to save files and e-mails to comply with record retention regulations. Energy firms, health care providers, educational institutions, and national and local government agencies, amongst other regulated industries, are all subject to mandates which require certain information to be kept for specified periods of time. With more of these records being created in electronic format (e-mail, online bank statements, etc), organizations have to implement electronic records management programs to meet regulatory requirements.

Some organizations, with the intent to improve corporate governance policies for the benefit of key stakeholders, are also using electronic records management programs to save unregulated, yet critical business data. Content such as board member communication, correspondence discussing executive compensation, and merger and acquisition contracts are being saved to help companies improve overall governance initiatives.

Failure to comply with record retention requirements can lead to fines and potential jail time, while failing to abide by corporate governance policies can create legal risks and other liabilities.

Business Intelligence

While compliance and electronic discovery force organizations to save more information for longer periods of time, business intelligence and knowledge management initiatives encourage information retention so that companies can leverage data as a corporate asset—historical information can help employees make more educated business decisions. For example, retaining all materials related to annual budget planning can help a finance department compare actual results against prognostications. By saving more information, finance can determine forecasting accuracy trends, improving future planning exercises.

The Common Requirement

Corporate data contains business records, potential electronic evidence, and content that employees can use on a regular basis. However, there is also a lot of other data that does not need to be saved at all. Organizations must find ways to determine what is worth keeping and, more importantly, for how long. Alternative approaches, including saving everything forever, are too expensive and can drive companies to forgo saving anything, which is also not feasible.

Organizations need solutions that automate the identification of archival candidates. This is not easy given that many companies can have several archive criteria, which can change over time. For example, an organization may archive all spreadsheets created by the mergers and acquisitions team. However, when this team engages in a transaction negotiation, a company may want to archive all content—e-mails, contract files, presentations, etc.—as part of a corporate governance policy. If an electronic discovery notice arrives, an entire finance department's e-mail may need to be archived, depending on the scope of the inquiry.

Depending on the company and the reason for archiving, policies may need to be based on the person creating the content, keywords within the data itself, when the information was created, and what type of file it is, amongst other criteria. To properly enforce these rules, companies need a way to define, dynamically update, and automatically execute within an archive solution.

Old Methods Won't Work

Backup versus Archiving

Organizations forced to retain digital information have traditionally relied on backup processes and technologies to do so. In fact, most IT departments do not differentiate between backup and archiving, with the latter commonly involving saving backup tapes for extended periods of time. However, with new accessibility requirements due to electronic discovery and business intelligence initiatives as well as the opportunity to store data in a separate lower cost storage infrastructure, organizations must now delineate between backup and archiving.

Backups are point in time copies of data stored separately from the primary instance of information. Most backup copies are created and saved in a compressed file format because the data sets are large. These backups can be leveraged to recover data in the event of deletion and corruption. In contrast, archived data is stored in native format after it has been indexed, making it easily searchable. In most implementations, archived data is not a copy of the primary environment; rather, it is a separate environment where data from the primary infrastructure is moved. The archive environment allows companies to enforce retention policies against a file or message as opposed to an entire backup data set, minimizing the amount of information ultimately saved.

Despite their differences, backup and archive processes can intersect. To make a backup of the primary infrastructure go faster, IT can first archive to reduce the amount of information to be copied. Additionally, because a corporate archive contains business records, potential electronic evidence, and other vital data, it should be backed up. Although it does not change frequently, customers should backup their archive to prevent against any data loss.

Manual Archiving

To their credit, some organizations have separated backup and archive processes for a variety of reasons. However, a portion of these companies rely on employees to move data into the archive. These employees need to know what data needs to be kept and for how long—an impossible task as records management, electronic discovery, and other factors driving retention policies change frequently.

Inconsistent execution is yet another problem plaguing manual archive practices—there is no way to ensure an employee is obeying retention policies. When information is being archived for storage management or business intelligence purposes, this may not be a significant problem. However, if a company is relying on employees to archive data for legal hold or regulatory compliance purposes, there can be substantial risks if the information is not saved or if the policies are not consistently enforced. A company may be accused of evidence spoliation or be unable to support or defend against a claim if data is not properly archived in an electronic discovery situation. For example, in 2007, Morgan Stanley was fined over \$27M for failing to properly preserve e-mail.² Intel went through a very public event when employees failed to preserve e-mails during an antitrust matter in March 2007.³

Forced Behavior

To date, many companies do not formally retain information at all—outside of backups—because of the associated storage costs. Instead, companies actually put into place various processes that force incoherent and inconsistent information retention practices. This occurs most frequently in e-mail environments as IT looks to control capacity, trying to reduce the storage costs and application performance issues that occur when the amount of messages and attachments being sent and received increases.

² Source: <http://www.informationliberation.com/?id=23908> (Posted October 1, 2007).

³ Source: <http://articles.latimes.com/2007/mar/08/business/fi-intel8> (Posted March 8, 2007).

Implementing a More Efficient Information Archiving and Retention Strategy

According to ESG research, two-thirds of North American organizations implement mailbox quotas.⁴ Employees have to create personal mail archives or constantly delete messages from their inboxes because if they go over their quota, they may not be able to send or receive any additional e-mails.

Aside from the productivity impact, e-mail quotas can cause other problems, especially during an electronic discovery request as corporate counsel or compliance officers must search through the primary environment as well as individual PCs and file shares where personal archives are stored. Additionally, if employees store personal archives on file shares, they still consume valuable storage capacity and take up space in the backup environment.

The Common Pitfall

Regardless of whether a company is forcing employees to create personal archives (mailbox quotas) or enable manual archiving, relying on employees to identify the right subset of content to be retained and actually save the information is an enormous risk. It leads to inconsistent archiving practices, which minimizes many of the benefits of retaining information in the first place.

Leveraging traditional backup processes for archiving creates the opposite issue: everything is saved for much longer than is typically required. Further, the data is saved in less accessible formats and is not easily searchable.

Companies must commit to an information archiving and retention strategy that is separate from existing methods of saving data. The burden of archiving must be shifted from employees to a solution that can consistently identify content to be saved, assign retention periods per file or message, and then store the messages in a searchable, more accessible format.

A Better Way

Purpose-Built Archive Software

An archive is a separate, tertiary environment where information can be managed and retained without impacting primary application infrastructure. By storing the information in a tertiary environment, customers can select what components—namely, servers and storage systems—make up that infrastructure. They can choose to leverage lower cost systems or purpose-built devices, such as WORM storage, to retain the information. With a tertiary environment, organizations can balance the cost to store data with accessibility requirements.

To move data into this environment, customers can leverage purpose built archive software solutions. Such solutions allow customers to set policies on what data is to be moved or 'archived,' execute the data movement, and then manage the retention once it is in the new environment.

During the archival process, the software typically indexes all the content, removes any duplicate files or messages, and assigns a retention period to the data. Employees can then search the archive based on keyword, date range, sender, or other criteria to find historical data. Corporate counsel may have 'super user' search permissions so they can query an entire archive environment, whereas individual employees can only access their specific messages or files.

The archive software is usually optimized to manage specific content types including e-mail, file systems, and SharePoint.

- E-mail archive solutions capture messages from the primary messaging environment via journaling or other methods and move them to a separate one supported by lower cost storage. With all the content

⁴ Source: ESG Research Report, *2007 E-mail Archiving Survey*, November 2007.

Implementing a More Efficient Information Archiving and Retention Strategy

in an archive, IT can still enforce mailbox quotas in the primary environment. When a quota is triggered, messages are simply moved out of the inbox to the archive and a stub or link is left in the inbox, facilitating seamless access.

With an e-mail archive, employees do not have to create personal archives to avoid mailbox quota related penalties. When a legal or regulatory discovery request arrives, attorneys have one central place to go to when search for relevant e-mails—as opposed to PCs, file shares, backup tapes, and anywhere else personal mail archives may be stored.

- Similar to e-mail archive solutions, file archive offerings can move content from one environment to another and allow customers to store a subset of information on lower cost storage. The file archive solution determines what content to move based on predefined policies. Employees can still access the archived files via stubs or links left in the primary environment or a web browser that provides an access to the central file (and e-mail if the customer has both implemented)archive.

The file archive also provides a single repository where employees—both individual productivity workers and corporate counsel—can quickly locate information.

- Microsoft Office SharePoint Server (MOSS) archive solutions are relatively new to the market, but operate in the same fashion as e-mail and file offerings. Customers define what data should be archived and the solution removes the object from MOSS and moves into a separate environment. The file can still be accessed via the MOSS interface.

What to Look For?

The good news for customers is that plenty of vendors offer purpose-built archive software solutions. The first challenge is finding one that makes it is easy to implement a cost effective information archive retention strategy. Before going into specific product features, customers should consider the following to help narrow down the list of vendors they may want to engage with:

- **Can the solution archive multiple content types with same platform?** You may want to implement an information archive and retention strategy across e-mail, file servers, and other content sources. Doing so with individual archive applications limits the benefits of having a centralized repository where retention policies can be consistently enforced and all data can be searched for a single interface.
- **Does the solution support for your application environment?** Some archive solutions do not support all messaging applications, while others may not have Microsoft SharePoint integration completed.
- **How does the solution scale in terms of capturing and processing messages?** An archive solution must be able to handle your company's content workload. This consideration applies mostly to e-mail—an archive solution must be able to ingest and process messages from any of the mailboxes you wish to archive. It must also be able to move data from file servers on a regular basis without impacting the performance of those devices.
- **How easy is it for employees to search and retrieve information?** It is somewhat pointless to archive data if employees cannot quickly query a repository and get relevant data back. Depending on the archive solution's architecture, search performance degrades as the amount of information archived increases. Additionally, companies that are archiving for discovery reasons should ensure an archive solution has hierarchical search roles that allow certain individuals to search the entire archive while employees can only see their content.

The second challenge is identifying an archive solution that addresses the aforementioned issues of automating the archive process itself as well as the management of archived information throughout the retention period. For customers that want to consistently enforce retention policies and cannot afford the resources to constantly manage the archive environment, they look for solutions that have the capability to:

Implementing a More Efficient Information Archiving and Retention Strategy

- **Identify content for archiving and set retention periods based on the file or message's metadata (file type, creation date, owner, etc.) or contents (keywords, etc.).** Customers may want to retain certain messages based on the sender, recipient, date, keywords within the body of the message, or a combination of any of these criteria amongst many other things. Having this flexibility helps customers archive only the data they want as opposed to saving non relevant business information.
- **Enable customers to dynamically update retention policies and archive criteria.** Records management, electronic discovery, and corporate retention policies can change frequently, so customers must have the ability to easily add or modify archive criteria and associated retention periods.
- **Allow customers to create their own metadata fields (or tags) so they can organize the archive based on their data and business requirements.** An organization may want to use its archived data as opposed to just saving it for the appropriate period of time. As an example, a company may archive anything created by the finance department and add a tag that delineates access to only those within that group. When a finance employee logs onto the archive, they can see all the content that is being saved by their peers. This makes an archive repository much more valuable as the information is grouped based on how it is used rather than why it is saved.
- **Take selective actions against individual items and messages during the retention period.** Once information is archived, it is still useful and must be managed as such. Data may need to be accessed as part of an electronic discovery request or companies may choose to lower storage costs by moving older data to a less expensive storage system. This management requires customers to take action with messages after they have been archived. Examples of managing an archive with selective actions are moving archived data to a different storage system; extending retention policies because criteria has changed, which often happens during and electronic discovery event (data needs to be saved longer than originally participation); saving archive searches so they can be run again when new content enters the archive; and deleting information if the retention period has expired.

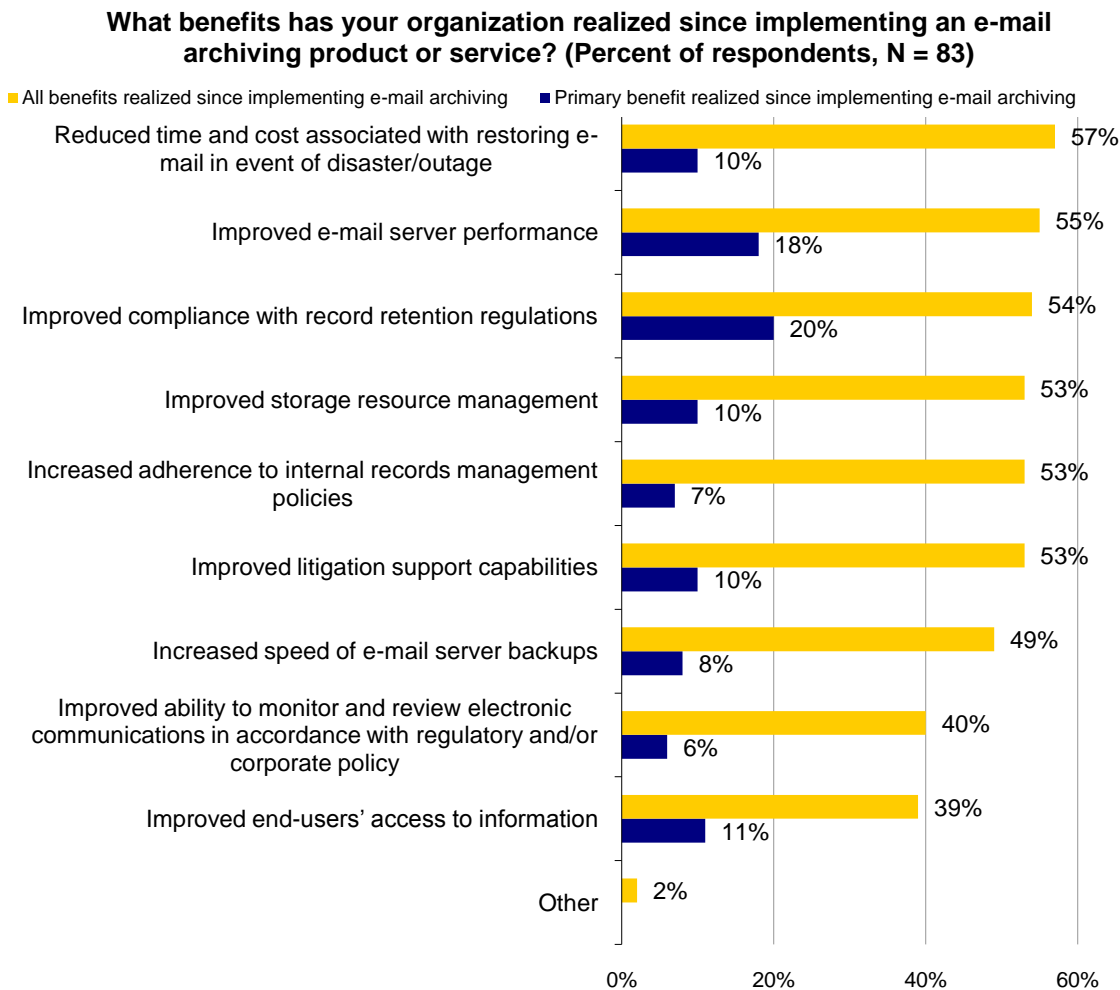
The Benefits

Regardless of which purpose-built archive solution an organization chooses to deploy, customers can achieve a number of benefits, including centralizing of information, which allows it to be easily searched and retrieved. This is especially helpful in electronic discovery situations when corporate counsels have limited time to identify, collect, preserve, and analyze information. Additionally, by moving the information out of the primary application environment, storage capacity is freed for newly created content. IT does not have to keep buying additional primary storage systems, which can save any organization a significant amount of money. Instead, IT purchases less expensive, denser devices to store the data in the archive.

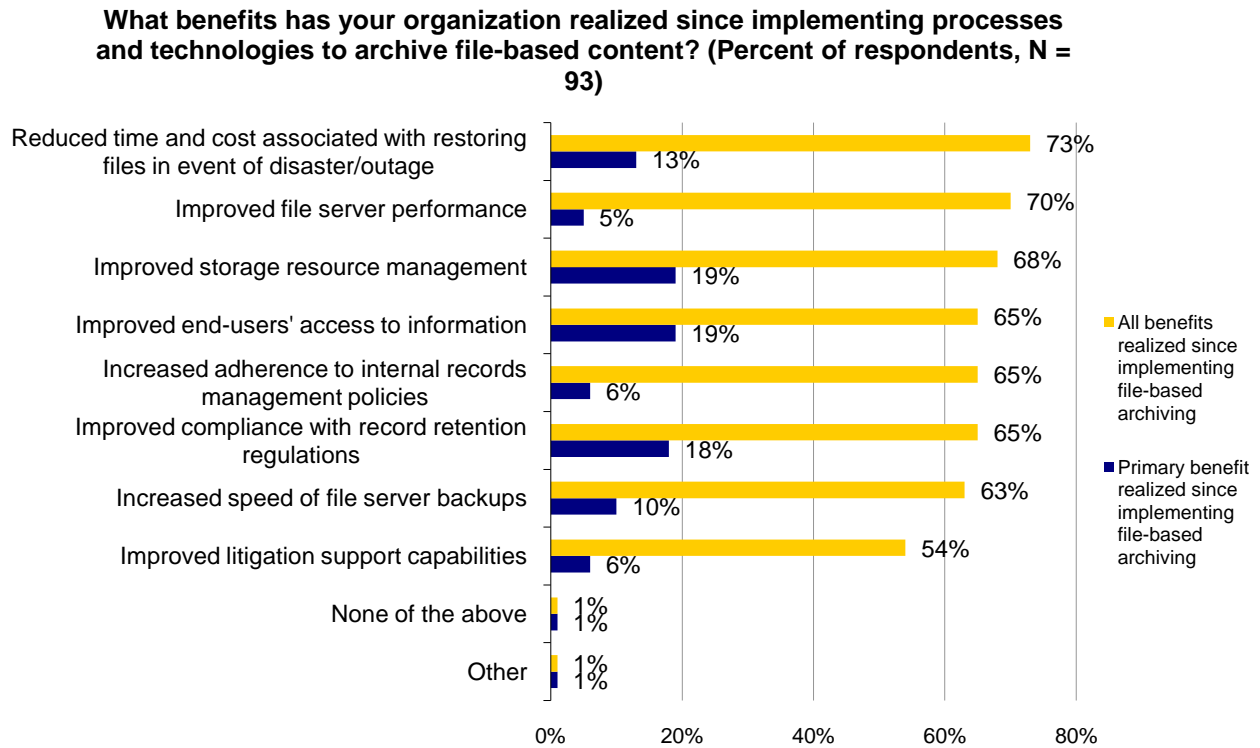
The movement of information out of primary environments should also boost primary application (or file server) performance. IT does not have to keep upgrading servers or buying expensive storage systems just to improve e-mail or SharePoint response times if data is constantly being offloaded to the archive environment.

Figures 3 and 4 highlight all of the benefits that current e-mail and file archive customers have experienced since their implementations. Because purpose-built archive solutions deliver so many benefits, many organizations can easily justify the investment and paybacks often occur much faster than originally anticipated.

FIGURE 2. E-MAIL ARCHIVING BENEFITS REPORTED BY CURRENT USERS



Source: ESG Research Report, 2007 E-mail Archiving Survey, November 2007.

FIGURE 3. FILE ARCHIVING BENEFITS REPORTED BY CURRENT USERS

Source: ESG Research Report, 2007 File Archiving Survey, December 2007.

The benefits of selecting the right purpose-built archive solution—specifically, one that automates archive and retention management tasks—can exponentially improve upon the above opportunities. For example, if an archive solution automatically archives content based on the creator, corporate counsel are ensured that all employee messages and files are retained as part of a legal hold process. As new data is created (or sent) by these individuals, it is kept according to the legal hold retention policies. As electronic discovery requests arrive, corporate counsel can simply update the archive criteria to capture the right information. Additionally, customers can extend retention policies on a per message or file basis if they are already in the archive and suddenly fall under the scope of the discovery request.

A purpose-built archive solution that can selectively take action—whether it is archiving or moving data once it is in the repository—can also drive down storage costs. Customers do not have to save everything. Companies can set and enforce retention periods on individual messages and files, enabling them to expire content once the retention period is up. Oftentimes, when a retention policy is set on a group of messages and one of the messages must be kept longer because of a legal hold, all of the messages in the group must also be kept, driving up storage costs. Further, as over half of the organizations surveyed by ESG said their e-mail archives are growing between 11-40% per year,⁵ companies need a way to store archived data more efficiently. Some purpose-built archive solutions enable organizations to move older data to lower cost systems. The information is still available through the archive application and customers are able to extend the concept of tiered storage to the archive environment.

⁵ Source: ESG Research Report, 2007 E-mail Archiving Survey, November 2007.

Conclusion

ESG expects overall corporate data to increase by an estimated 25% in 2009 after several of years of growth at two to three times that rate.⁶ Combine this data growth with all of catalysts to retain information—corporate governance, electronic discovery, compliance, and knowledge management initiatives—and it is easy to see why companies need to improve overall information archive and retention strategies as well as the technologies that support those strategies.

It is not easy to make the leap from mailbox quotas and tape-based archives to purpose-built archive solutions. However, users will soon realize that setting up a separate archive environment provides more control—ensuring data is actually saved and improving information accessibility. The key for organizations that have yet to make an investment in one of these solutions is to really think about the features that make it possible to save the right data for the appropriate length of time and manage the archive as an active, useable source of information. Doing so will lead to substantial measurable benefits, including dramatically cutting both primary and archive-specific storage costs while facilitating more efficient electronic discovery operations—both of which should help deliver a very fast, tangible return on investment on purpose-built e-mail archive solution purchase.



20 Asylum Street
Milford, MA 01757
Tel: 508-482-0188
Fax: 508-482-0218

www.enterprisestrategygroup.com

⁶ Source: ESG Research Report, *Enterprise Storage Survey*, November 2008.