Osterman Research

WHITE PAPER

White Paper by Osterman Research
Published Date **December 2018**
Sponsored by Zix

Why You Must Archive Business Content and What You Can Do With It

Executive Summary

Among organizations that archive their email and other electronic content for long periods, the typical practice that they employ is "Write Once Read Never": they capture, index, store and manage this content, but then rarely access it again unless it's needed to support an eDiscovery effort, a regulatory compliance audit, or if a user needs to retrieve a missing file. And that's not necessarily a bad thing: organizations should archive their business records for these kinds of defensive purposes, since there will inevitably be a lawsuit, a regulatory requirement or some other reason to go back into the archives to extract information that is needed only occasionally to satisfy a request.

However, there are other reasons to deploy and maintain an electronic archiving system in combination with a robust analytics capability that can mine for insights from the archived data. Decision makers should consider what they can do with their electronic content proactively, such as extracting business intelligence from the stored data, finding problems in the organization that might lead to a data breach, and better understanding the sales process and the actions that lead to higher customer retention rates, among many other things.

In short, the right solution for archiving electronic content provides a foundation for supporting not only the conventional, defensive applications for data retention; but also the proactive applications that can enable competitive and other advantages.

Decision makers should consider what they can do with their electronic content proactively, such as extracting business intelligence from the stored data.

KEY TAKEAWAYS

- Today, most organizations retain corporate email and users' files as part of their archiving process, but most do not archive other content types.
- There are a number of reasons for deploying an archiving solution, but the most important reasons are compliance with regulatory obligations (including privacy obligations), eDiscovery and litigation holds.
- There are other reasons to deploy an archiving solution, including storage management, maintaining a record of corporate history, and giving users the ability to search for older emails and other content.
- While mostly overlooked in the past, extracting insight and intelligence from archived data is becoming a major component in archiving ROI. This focus of archiving will become much more important over the next two years.
- The ability to extract insight and intelligence from corporate archives offers a number of important benefits. Use cases include improved customer service, prospect management, more readily finding insider threats, discovering employee violations of corporate policy or best practice, conducting investigations, and understanding employee sentiment and behavior.

ABOUT THIS WHITE PAPER

This white paper program included an in-depth survey of current and planned archiving practices by mid-sized and large organizations, some data from which is included in this paper. However, a full survey report will be provided in a separate document shortly after this paper's publication.

This paper was sponsored by Zix. Information about the company is included at the end of this paper.

Why Archive Electronic Content?

Electronic archiving is a long-standing practice, particularly in heavily regulated industries like financial services and healthcare, in which regulators have required industry participants to retain business records for long periods. However, long-term data retention is a requirement across a wide range of industries – the US federal government, for example, has imposed data retention requirements across just about every industry for many different types of records.

What these data retention requirements have in common is that they are almost exclusively defensive in nature – organizations must retain data for long periods to satisfy their legal and regulatory obligations. However, there are other reasons to retain data, but these also focus on a more defensive justification – protecting against data loss or user mistakes, for example.

THE TRADITIONAL DRIVERS FOR ARCHIVING

The primary, traditional reasons for archiving electronic content are driven by a number of considerations:

- **Legal obligations**
  Just about every organization is subject to a variety of legal and contractual requirements. As a result, they need to retain various types of electronic content in the event this content is needed in the future to support their role as a defendant, plaintiff or third-party participant in legal proceedings. The requirement to retain and manage data is imposed from a variety of sources, including legal precedent (courts establish standards for the length of time that data must be retained), statutory obligations (specifically defining the retention and production obligations for certain types of data), and internal best practices. Plus, organizations that reasonably expect pending litigation, such as a wrongful termination lawsuit, may also need to subject some electronic content to a legal hold period that is different from their standard policies. A centralized archive can facilitate that process in way that traditional backups cannot.

  Organizations that manage their eDiscovery capabilities using a centralized and properly maintained archive are normally much better off than if they rely solely on backups to do so – information is easier to find, the process is faster, the data set they have captured is more complete, and it's usually much less expensive.

  The ability to search and access electronic records, especially across the various siloes in which an organization's data is typically stored, can permit legal counsel to understand a case before investing substantial time, money and effort in electronic records retrieval.

- **Regulatory compliance**
  Many electronic records that relate to an organization's business activities are subject to a variety of regulatory compliance obligations. These vary widely by industry and jurisdiction. Decision makers should be mindful that virtually every organization and industry faces some type of regulatory compliance requirements to retain records, and these retention obligations are not limited to just "regulated" organizations or industries.

  Retention obligations include the retention of content like employee records, financial documents, email correspondence, invoices, shipping information and a variety of other files and content types. In fact, even metadata must be preserved.

  Although many industries have significant numbers of retention obligations, among the more heavily regulated verticals around the world is the financial services industry. In the United States, for example, rules of the Securities and Exchange Commission (SEC) and the Financial Industry Regulatory Authority

Many electronic records that relate to an organization's business activities are subject to a variety of regulatory compliance obligations.

(FINRA) require members of national securities exchanges, brokers and dealers to retain securities transaction records for at least six years. In Canada, records of purchase and sell orders of securities must be retained for seven years. In the United Kingdom, investment service and transaction records must be retained for at least five years. A failure to comply with these retention regulations can be severe and often involve the imposition of significant financial penalties.

Healthcare is also a heavily regulated industry. For example, the Health Insurance Portability and Accountability Act of 1996 (HIPAA), obligates organizations to protect patients' electronic health records from unauthorized users and to retain this information for six years. Non-compliance with HIPAA requirements could result in fines of up to $50,000 per violation, or criminal penalties of $250,000 and up to 10 years in prison for the most egregious violations.

- **Privacy regulations**
  There is a growing number of privacy regulations being initiated around the world. Although the European Union's General Data Protection Regulation (GDPR) is the most prominent of these regulations, there are regulations either implemented or soon to be implemented in California, Colorado, Brazil, Australia, Japan and many other countries.

  The importance of archiving as an enabler for compliance with privacy regulations cannot be overstated. Using the GDPR as an example:

  - Article 15 gives data subjects the right to ask any entity that possesses or processes his or her personal data (a data controller) to produce that data on demand. Without an archiving solution that enables a search across all information that a company controls or processes, compliance with such a request would be extremely time-consuming, if even possible.
  - Article 30 requires that data controllers keep records of their data processing activities, with a list of specific information to be retained for each record. Here again, archiving is essential to ensure that all such records are captured and later accessible.
  - Article 17 states that, subject to certain conditions, a data subject has the "right to be forgotten" by any data controller that possesses or controls his or her information. An archiving solution can ensure that the information knows the information it has on each data subject and can defensibly delete it when required to do so.
  - Moreover, adequate controls are required for copies of production databases that contain personal data taken for testing, development, or analytics purposes; data sources that contain customer contact and profile data; and many other types of data.

The importance of archiving as an enabler for compliance with privacy regulations cannot be overstated.

- **Storage management**
  An archiving solution can improve system performance by minimizing the amount of "live" data that must be stored on active servers. Because electronic data like older email messages and files are accessed infrequently, it often makes sense to move this content to an archiving system for better system performance. This can minimize the amount of time required to backup email and data servers, and it can speed the time to restore a server from backups.

  Another important benefit of archiving in the context of storage management is the ability to deduplicate and provide single-instance storage, which offers the potential to dramatically reduce storage requirements.

- **End-user self-service**
  In order to satisfy employees' need for business information, email, collaboration

tools and other electronic content repositories are often relied upon as the primary tools used for doing work. For example, an employee may need to locate older emails quickly so that he or she can review their own email correspondence or other content, such as attachments, in email. Alternatively, a new employee may have to trace back email and other electronic content between his or her predecessor and a customer.

- **Knowledge management and retention of corporate history**
  Email and other electronic content are typically one of an organization's most important sources of corporate knowledge. Some analysts have estimated that most of an organization's intellectual property is housed in its messaging systems. Even if that is overstated, an organization's electronic content does contain important (structured and unstructured), employee-generated information necessary for its growth.

WHAT DO ORGANIZATIONS ARCHIVE AND WHY?

At present, most organizations retain corporate email and users' files, but not much else, as shown in Figure 1.

Figure 1
Percentage of Organizations Retaining Various Content Types

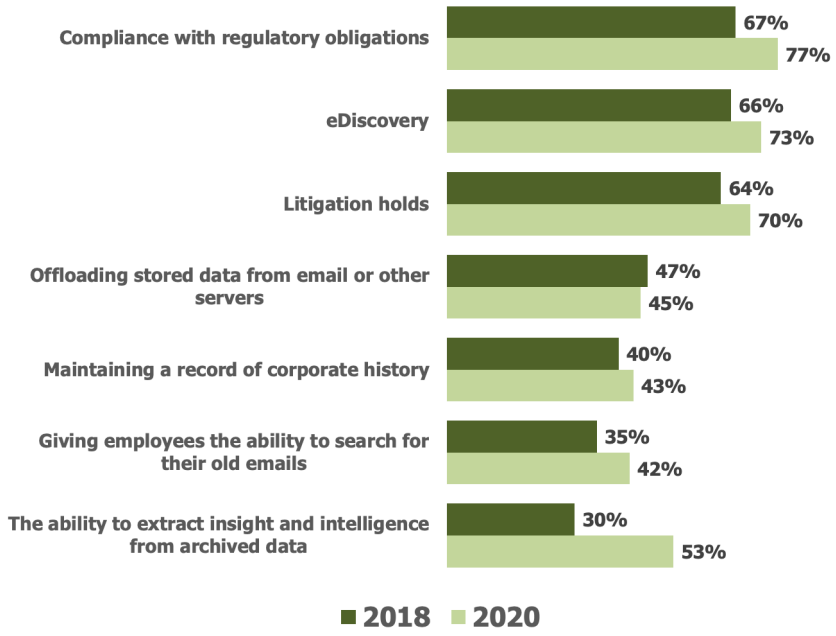
Content Type	%
Corporate email	93%
Users' files	62%
Content from SharePoint or similar collaboration tools	41%
Content from corporate collaboration solutions (e.g., Slack, Teams, etc.)	33%
Company website and blog content	28%
Voicemails from the company phone system	27%
Content from company-managed file sync and share tools, e.g., Dropbox	26%
Company videos	20%
Work-related content from employees' instant messaging accounts	16%
Content from company-owned mobile devices	16%
Corporate social media posts	15%
Content from users' personally managed file sync and share tools	15%
Work-related content from employees' personal mobile devices	11%
Work-related posts from employees' personal social media accounts	8%
Other	9%

Source: Osterman Research, Inc.

Most of an organization's intellectual property is housed in its messaging systems.

The research conducted for this white paper found that the traditional drivers for archiving – regulatory compliance, eDiscovery and litigation holds – dominate most organizations' reasons for deploying an archiving system. However, over the next two years, while these drivers will become somewhat more important, the fastest growing reason for deploying archiving solutions will be the ability to extract insight and intelligence from archived content, as shown in Figure 2.

Figure 2
Drivers for Deploying an Archiving Solution, 2018 and 2020
Percentage Responding an "Important" or "Major Driver"



Source: Osterman Research, Inc.

Should You Be Proactive?

The "defensive" reasons for archiving electronic content are well-established and fairly straightforward. But how about the "offensive", or proactive, reasons to retain this content?

WHAT DOES AN ARCHIVE INCLUDE?

An archiving solution focused solely on archiving email, for example, contains a number of different types of information, including:

- Every relevant email sent between employees (by contrast, irrelevant emails that typically do not need to be archived are of the "birthday cake in the conference room at 3:00pm" variety)
- Every relevant email sent to or from every employee outside of the organization
- All attachments that flow through the email system
- The date stamp for every email sent and received
- The wording and tone of words in each email
- Whether or not sensitive data was properly encrypted before it was sent
- The timing between receipt of and response to each email
- Which employees are working after-hours on company business
- If employees are using corporate email for personal uses
- Whether or not emails requesting a response actually received one, and

The "defensive" reasons for archiving electronic content are well-established and fairly straight-forward. But how about the …proactive, reasons to retain this content?

- Various types of metadata associated with each email.

If an archiving solution is used for additional data types, such as social media posts, instant messages or text messages, other types of information can be retained, such as social media posters' views about clients and competitors, their views on other employees, their use of profanity in non-email channels, their use of non-email systems for sending corporate content, and the like.

WHAT ARE SOME OF THE USE CASES?

There are a number of use cases for an archiving solution in combination with a powerful analytics capability; here are just a few examples:

- **Customer and prospect management**
  The content and timing of every customer's or prospects' inquiry, complaint, request for more information, etc. can be tracked. Data on each response can also be tracked, including how long the response took, who provided the response, the tone of the response, whether or not the customer responded, etc. This archived data can provide significantly more information than might be available in some CRM systems, since the archiving solution automatically tracks all of this data.

  This data can then be used to determine if there is a correlation between the length of time it takes to respond to a prospect's inquiry and the likelihood of making a sale, or if there is a relationship between customer renewal rates and how quickly their complaints are addressed. This information will help decision makers understand how to modify the customer management or prospecting process, determine who the best performers might be, or how to triage customer complaints in the most efficient manner.

- **Finding likely insider threats**
  The tone and content of managers' communications to their employees can be monitored for problematic behavior. For example, employees who are berated by their managers are more likely to steal data or finances, and so examining archived data can help senior decision makers to find and deal with problem managers before an insider threat can occur.

- **Reduced use of profanity can be an indicator of wrongdoing**
  IBM has developed an analytics capability for monitoring traders for potential signs of wrongdoing. In the United States, IBM has found that traders who reduce their use of profanity may be up to no good[1]. Interestingly, just the opposite is true in the United Kingdom – traders who increase their use of profanity may be indicating that they are involved in malicious activity.

- **Heading off embarrassing situations**
  Employees' comments about clients or managers can be monitored to determine if there are issues that should be addressed in a timely manner. For example, the following tweets were gleaned from Twitter in mid-November 2018 along with the identity of the individual posting them that can easily be traced back to their employer:

  - "My boss got banned from Applebee's a while ago for always breaking glasses when him and his friends got drunk and I still find it hilarious to this day"

  - "First night in La Crosse for work, got drunk with a bunch of tech theatre teachers.

There are a number of use cases for an archiving solution in combination with a powerful analytics capability.

[1] https://www.ibtimes.com/wells-fargo-scandal-banks-tap-watson-monitor-employee-activity-2586425

- "On my 2nd week at my new job, I lied."
- "Just got a email about a mandatory company wide conference call... A call to tell us our company is being sold and new owners are prob closing us"

Few managers would want this type of information being relayed to new customers or prospects. An archiving solution can be used to search for this type of information so that corrective action can be taken or corporate policies updated.

- **Detecting policy violations**
  The use of personal webmail to conduct company business can be tracked either directly or indirectly by searching through archived content to identify violations of corporate policies against use of personal resources to conduct company business.

- **Identifying and tracking real world communications**
  A comprehensive solution that archives email, social media, text messaging and other content and channels can be used to identify and track how "real world" communications takes place. For example, a sales rep may initiate a business communication via text messaging, follow up with an email, and then send a file using personal webmail when the corporate email system goes down. Archiving the gamut of potential communication channels is essential to capture all communications with an organization.

- **Conducting investigations**
  Conducting investigations is a key capability for a robust archiving solution. However, a difficult challenge is getting a clear understanding of what took place and when it occurred. Too often, people do not have a clear recollection of what took place or what might have been said. Fortunately, email, text messages, instant messages and other content provide clear evidence of what was said (assuming that chain-of-custody was preserved).

- **Understanding employee sentiment and behavior**
  A robust archiving solution with good analytics can identify problems so that violations of corporate policy, the law, or best practice can be addressed before they result in a more serious problem. For example, a company's compliance staff could search for evidence of sexual harassment, illegal downloads, distribution of offensive content, or any of several other activities that might result in a lawsuit, regulatory action, scandal or some other issue.

The use of personal webmail to conduct company business can be tracked either directly or indirectly.

It's important to note that a robust archiving-plus-analytics capability does not need to include every communication type used in an organization, since the vast majority of communications in most organizations occurs in email. However, adding additional content types like social media posts can enable additional insights and corporate intelligence to be extracted from the archive.

Many Solutions Were Not Designed for the Next Generation of Archiving

Early generation email archiving solutions were designed with a focus on managing mailbox size. In the early days of email, mailbox size was limited to only tens or hundreds of megabytes. These solutions were designed to remove email and attachments that were consuming a significant amount of storage and replace them with a small "pointer" or "stub" to the archive. This feature allowed users to keep months' worth of email in their inbox without exceeding the mailbox size limit.

Today, more modern email solutions like Office 365 support multi-gigabyte mailboxes, capable of holding orders of magnitude more data than older solutions.

Consequently, mailbox size management is much less of a driver for email archiving than it used to be and we see its importance in the context of archiving continuing to decline. That doesn't mean that users no longer need archiving for mailbox management, because some users continue to run into mailbox size limits – this is especially true for users who employ email as their primary file-sharing solution instead of using file-sharing technologies like Microsoft OneDrive or Box.

Many organizations use journaling to retain electronic content. Journaling retains a copy of all email that sent and received for each mailbox. It is the responsibility of the archiving solution to protect the journal email copy. While in Office 365 environments, for example, an Exchange Online mailbox cannot be designated as a journaling mailbox, for organizations that run an Exchange hybrid deployment with mailboxes split between on-premises servers and Office 365, administrators can designate an on-premises mailbox as the journaling mailbox for Exchange Online and on-premises mailboxes.

Among the problems with many conventional archiving solutions are:

- **Data is locked away in siloes**
  Electronic content is normally stored in a number of independent silos, such as email, mobile carriers or CRM. The growth of corporate applications, particularly those in the cloud, and IT's growing acceptance of the "Bring Your Own" trend for applications and devices, means that information management is becoming increasingly fragmented and distributed. This makes data more difficult to access by senior managers, legal teams, compliance and others who must access it. Ad hoc data queries, such as those that individual employees might initiate to look for their older content, are more difficult if an organization is storing a variety of data types in disconnected siloes. The result for end users is that they often do not look for older data because of the problems involved in doing so. Instead, they will recreate what they need, reducing employee productivity in the process.

- **Data is hard to access**
  Information is increasingly distributed and disconnected, and this leads to an inability to search and synthesize data easily across the various siloes in which it is stored. This means that information access requires visiting a number of data siloes one at a time, such as email, CRM, social media, etc., each with its own interface. Plus, data cannot be connected between siloes in most cases.

- **Most solutions were not designed for analytics**
  There are a large number of archiving solutions in place today that just were never designed for robust analytics. Many of these solutions do a good job at capturing, indexing and retaining data – and many can do so at scale – but they were not designed to provide the level of in-depth analytics that next-generation archiving demands.

There are a large number of archiving solutions in place today that just were never designed for robust analytics.

Some Recommendations

Any set of recommendations for moving forward with a next-generation archiving approach will be dependent upon a number of factors, but Osterman Research offers the following, high-level recommendations for consideration:

1. **Deploy an archiving solution**
   The traditional role of archiving is primarily a defensive one, but a critically necessary one. Organizations faced with regulatory or legal obligations to retain and produce data must have a robust and scalable archiving solution. Even so, many organizations still don't archive their content or they do so in a haphazard way. We recommend that non-archiving organizations look at email archiving as a first step. Cloud-based enterprise email solutions, e.g., Microsoft Office 365, offer built-in archiving or archiving options, and so a separate, third party email solution is not a requirement in all situations. While Osterman Research highly

recommends the use of third-party archiving solutions in most cases, they are not an absolute requirement to get started with a basic archiving capability.

2. **Decide how information should be retained**
   Decision makers should establish policies describing how all of their data should be retained and managed. One option is to collect copies of archived data into a central repository and maintain it under IT control in a single archive. While this can be a viable solution in many cases, it can cause some problems. For example, there could be an increase in storage requirements because the central archive will now contain a duplicate copy of archived content that is also stored elsewhere. At least some of the data in the central archive will be temporarily or permanently out of sync with data in the original archives as changes are made to the latter between replication/syncing cycles. Plus, not every piece of content needs to be archived, and so copying all data to centralized archives will store large quantities of data unnecessarily. This adds not only to storage and storage management costs, but also to legal and compliance costs when the data is searched.

   A better option may be to implement a solution that will enable retention and analysis of content "in-place" instead of moving it into a centralized archive. This can reduce storage requirements, can make data management much easier, and allow easier integration of legacy data stores into the overall content repository.

3. **Focus on extracting insight and intelligence from corporate data**
   Extracting information from archived content is valuable and can provide insight available from nowhere else. Performing analytics on this data can offer insights to all levels of an organization and may permit them to gain new insights from their archives. However, decision makers must understand these benefits and authorize the resources necessary to realize them. Getting buy-in for these kinds of solutions is not an easy thing to do in some organizations, especially in those that are not yet sold even on the concept of traditional archiving.

4. **Sell the use cases**
   We presented several use cases in this paper of how next-generation archiving can help drive better decision-making and enable greater employee productivity, but these merely scratch the surface of what's possible. There are a large number of potential use cases that will be unique to individual organizations and that can provide competitive advantage, reduce risk and otherwise add value.

Decision makers should establish policies describing how all of their data should be retained and managed.

Summary

Defensive archiving – the practice of retaining all relevant business records for the appropriate length of time for legal, regulatory and compliance purposes – is an essential best practice for any organization. However, combining a robust archiving solution with an analytics capability is increasingly becoming a best practice because it can enable decision makers to glean insights and intelligence from archived data for purposes of enabling competitive and other advantages.

Sponsor of This White Paper

To better meet your company's security, data protection and compliance needs, Zix can enhance your Office 365 environment with advanced threat protection, archiving and email encryption. Zix delivers a superior experience and easy-to-use solutions that have earned the trust of more than 19,000 organizations including the nation's most influential institutions in healthcare, finance and government.

To defend your company from malware, ransomware, phishing and other email threats, ZixProtect combines a multi-layer email security approach with automated traffic analysis, machine learning and real-time threat analysts. In addition, ZixProtect's business continuity feature ensures that your organization can continue to communicate if your email experiences a disruption.

ZixArchive eases email archiving and eDiscovery with automatic email collection and storage in a secure cloud. Its automatic indexing and multiple search criteria gives you and your employees convenient and rapid access to archived emails. ZixArchive also enables you to share an email hold with outside legal counsel and auditors and revoke privileges when access is no longer needed, keeping your data within your control.

To ease email encryption for you, your employees and your recipients, leverage the industry's leading solution ZixEncrypt. Automatic transparent delivery between customers and robust delivery methods for other recipients enables easy access to encrypted email for anyone, anywhere and on any device, making the user experience exceptional and compliance simpler. Proven policies and advanced reporting provide peace of mind, while customizable branding and security capabilities make email encryption fit your unique company needs.

Leveraging our more than 15 years of hosted experience, you can have confidence that Zix email security solutions integrate seamlessly with Office 365. You also benefit from the support of the ZixData Center, a state-of-the-art facility with PCI DSS 3.2 certification, SOC2 accreditation and SOC3 certification. Staffed 24/7/365, ZixData Center has a track record of consistent 99.999% availability. In addition, Zix delivers exceptional customer support 24/7/365 no matter your questions or concerns. With reliability, experience and superior support, Zix improves email security for your Office 365 environment. To learn more about our solutions for Office 365, visit www.zixcorp.com/office365.

www.zixcorp.com

@ZixCorp

+1 866 257 4949

sales@zixcorp.com