WHITE PAPER

# Next-Generation Archiving: Extracting Intelligence From Your Electronic Archives

**An Osterman Research White Paper**



**Osterman Research, Inc.**
P.O. Box 1058 • Black Diamond, Washington • 98010-1058 • USA
Tel: +1 253 630 5839 • Fax: +1 253 458 0934 • info@ostermanresearch.com
www.ostermanresearch.com • twitter.com/mosterman

# EXECUTIVE SUMMARY

What can your organization do with an archiving solution – an on-premises or cloud-based system that will ingest email and other electronic content, index this information, place it into archival storage, and allow the search and production of relevant content from that archive?

The traditional role of archiving solutions – which have been deployed in about 70% of mid-sized and large organizations[1] – is primarily defensive in nature: to retain information in the event that a litigation hold must be implemented, to be able to produce information in response to a court's discovery order, or to satisfy a regulator's demand for the production of relevant data. Other traditional applications of archiving include helping IT to more effectively manage storage on email and other servers, to enable users to find their older electronic content in a self-service mode instead of relying upon IT to do this for them, and to retain information for purposes of preserving corporate "memory".

## NEXT-GENERATION ARCHIVING

However, an email or other information archive contains enormous quantities of meaningful data about a variety of things in an organization:

- The platforms employees use to communicate
- How communication modes differ internally and externally
- The frequency of their communication
- How quickly salespeople respond to customers and prospects
- How well (or not) employees are following corporate compliance guidelines
- How managers are treating employees
- The social relationships that facilitate problem solving
- Employee sentiment analysis

In short, archives contain a wealth of information about an organization – data that can be analyzed in order to gain intelligence about an organization and deep insight into its operations. Consequently, instead of deploying and maintaining an archive solely for defensive purposes, next-generation archiving solutions will be deployed to mine meaningful data from archived information in order to help decision makers learn more about their organizations.

## KEY TAKEAWAYS

The key takeaways in this white paper are:

- Of the various motivating factors for deploying an archiving solution, the ability to extract insight and intelligence from archived data will see the greatest increase in interest among decision makers over the next two years.

- Traditional drivers for archiving – focused primarily on regulatory compliance and litigation support – will continue to be important, but next-generation archiving will be an important driver for deployment of new archiving solutions. In addition, next-generation archiving will enable compliance departments to get ahead of litigation and regulatory issues by proactively monitoring interactions for potential issues before they are surfaced by external entities.

- There are a large number of potential use cases for next-generation archiving solutions, but an expanding pool of decision makers outside of IT will need to be educated about their value. As the business value of intelligent insights increases, these use cases for next-generation archiving will expand beyond IT, requiring vendors to help coach businesses about how to best position and leverage these solutions in the organization.

[1] Source: *Survey on Email Systems and Next-Generation Archiving*, Osterman Research, Inc.

- Most current archiving solutions are not designed to be used as next-generation archiving systems and will need to be either supplemented with new capabilities or replaced. The survey conducted for this white paper found that only 10% of organizational decision makers and influencers believe their current archiving solution is capable of supporting next-generation capabilities.

### ABOUT THIS WHITE PAPER

This white paper discusses the benefits of next-generation archiving, some of the results of a survey that was conducted for this paper, and an overview of Micro Focus, the paper's sponsor.

## THE TRADITIONAL DRIVERS FOR ARCHIVING

An electronic archiving solution ingests email and other digital content, indexes that content, and places it into archival storage where it can be accessed using robust search tools. Archiving has been a common best practice for the last couple of decades, first in heavily regulated industries like financial services, then across all industries and organization sizes. This "traditional" view of archiving has been a defensive one: namely, retaining and protecting information in the event that it must be produced to defend the organization in some capacity. Consequently, the primary drivers for archiving have been and continue to be:

- **Litigation support**
  Virtually every business organization will eventually become a party to a legal action in some capacity: as a plaintiff, as a defendant, or as a third party. As a result, these organizations must preserve information they might need in order to pursue a legal action or defend themselves in one. Every organization has an affirmative duty under the Federal Rules of Civil Procedure (FRCP) to search for and produce relevant electronic content, such as email. This duty normally attaches when a party knows or should reasonably know that its electronic content may be relevant in a legal case. The duty to produce this information might be based on an eDiscovery order handed down by a court, or it can arise when decision makers become aware that litigation might be upcoming, requiring the organization to place a "litigation hold" on relevant data in order to prevent its deletion.

- **Regulatory compliance**
  Every organization has some level of regulatory obligation to preserve and monitor content. While the strictest regulatory obligations are placed on a handful of industries, such as financial services, healthcare, government, energy and life sciences, every organization has an obligation to retain certain types of data. These regulations require an organization to retain content like financial documents, certain types of email correspondence, employee records, and some types of client records. Even metadata must be preserved in many cases.

- **End-user self service**
  One of the primary benefits of an electronic archive is the ability for end users to access their older content for long periods without the requirement to store this data on "live" servers or in an email inbox. By archiving older content and giving users access to their own archive, both IT and end users benefit: IT can place strict quotas on mailbox size, for example, which speeds the backup and restoration of servers; and users benefit by having access to content as old as they need. Moreover, users can readily access their older content without having to bother IT with requests to search for and restore these files.

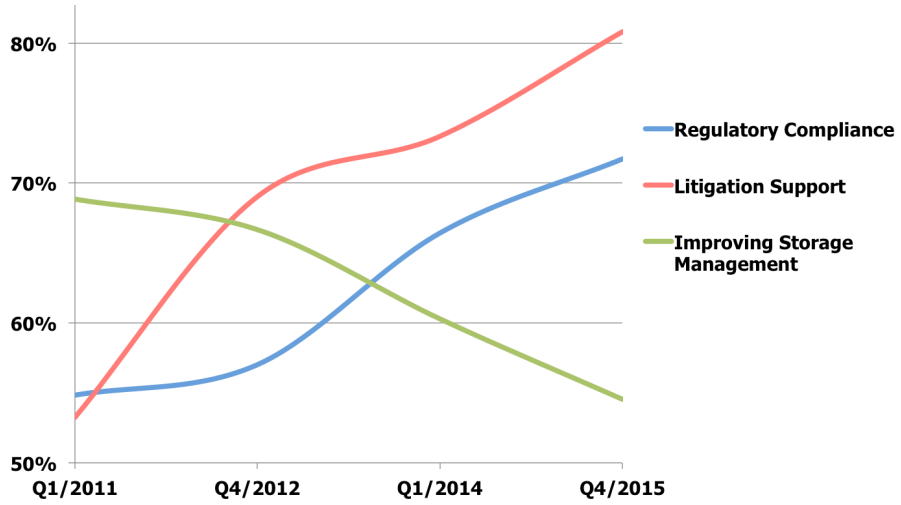
- **Storage management**
  Storage management—primarily for on-premises archiving solutions—used to be a primary consideration for deployment, since an archiving solution could absorb older content without the need to maintain it on first-tier storage. While storage management remains an important consideration, the decreasing price of

storage, the resulting growth in mailbox sizes and migration to cloud archiving mean that storage management is becoming less important over time as a driver for traditional archiving.

- **Retention of corporate history**
  Another important driver for archiving is the ability to preserve relevant information to maintain an appropriate record of an organization's history. However, the use of an archiving solution can also to help minimize problems that could arise in the absence of such a record.

The primary drivers for traditional archiving, and how these have been changing over time, are shown in Figure 1.

**Figure 1**
**Drivers for Electronic Archiving, 2011-2015**
*% Responding an Important or Extremely Important Consideration*




***Source: Osterman Research, Inc.***

A March 2016 Osterman Research survey found that the ability to extract insight and intelligence from archived data is currently only a minor driver for most organizations today: among IT decision makers and influencers, only 24% considered it to be a "driver" or "major driver", making it the least important motivator for organizations to deploy an archiving solution. However, when asked how they would view this application of archiving in just two years, 33% indicated that extracting insight and intelligence from archived data would be a driver or major driver, making it the fastest growing motivator for deploying an archiving solution. As non-IT stakeholders are made increasingly aware of the benefits of next-generation archiving for extracting insight about their customers, sales processes, employee sentiment and other applications, the demand for next-generation archiving will accelerate.

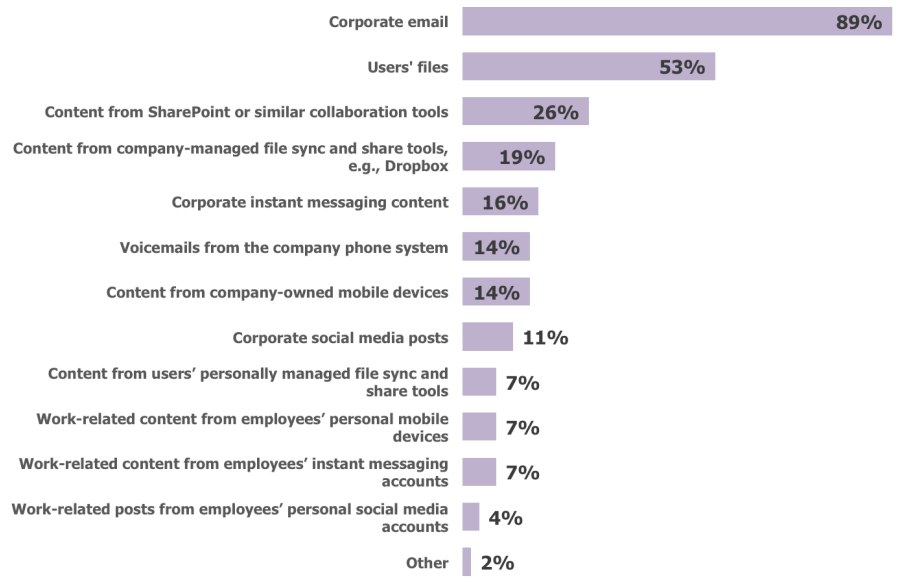
## THE NEW DRIVER FOR ARCHIVING

Electronic content archiving is entering its second phase – what we're calling Archiving 2.0, or Next-Generation Archiving. While the defensive role of archiving is critical and an important best practice for every organization, decision makers need to consider archiving's vital role in extracting meaningful business intelligence and additional value from the electronic content the organization already possesses.

## WHAT DOES AN ARCHIVE CONTAIN?

Electronic archives contain a great deal of information. For example, a January 2016 Osterman Research survey of end users[2] found that the typical information worker sends a median of 30 emails per day and receives 100. That means that during a 250-day workyear, each information worker sends or receives a total of 32,500 emails. An organization of just 500 information workers, therefore, will generate more than 16 million emails every year. This is in addition to employees' social media posts, instant messaging conversations, collaborative work, telephone calls and other communications both internally and with business partners, customers, prospects and others. As shown in Figure 2, the typical corporate archive contains a wealth of information.

**Figure 2**
**Types of Electronic Content That Organizations Archive**
*% of Organizations Archiving This Content*




***Source: Osterman Research, Inc.***

## WHAT'S IN AN ARCHIVE?

What does archived content include? Considering email alone as a source of information, an archive may include many things, such as:

- A second-by-second record of what information workers are doing during working hours and what they do when they are away from the office working on company business.
- A record of every communication between employees, customers, prospects, business partners, suppliers, contractors and others.
- How employees deal with sensitive content, such as sending or storing content via corporate email, via their personal email systems, in file sync and share tools, or via text messaging.
- A record of how employees respond to customers and prospects, as well as those communications to which employees did not respond.

[2] Source: *Results of a Survey of End Users' Messaging, BYOD and Social Media Practices*

- A record of every business relationship maintained by employees.
- A timeline of employees' responses to time-sensitive communications.
- A record of every appointment set by employees, their location at various times, and other relevant information about how they spend their day.
- The communications that employees have with competitors.

A detailed analysis of email content will reveal a wealth of information that can help decision makers to analyze their business operations more thoroughly. For example, performing analytics on email content will reveal:

- The extent to which employees comply with corporate policies about protecting sensitive information or comply with email acceptable use policies.
- Overall corporate sentiment based on employees' and management's communications.
- How salespeople interact with customers and prospects, how quickly salespeople respond to customer inquiries, and their tone in these responses.
- How well the official organization chart matches up with the actual one. For example, understanding email flows within an organization can help identify those individuals who wield the greatest influence in an organization – the "mayors" to whom employees go to solve problems and get things done.
- The social relationships existing within a company that can help fellow employees find the expertise and influencers they need to work more effectively.

Moreover, it should be noted that while email is the primary method for communications, collaboration and file transfer in most organizations, it is certainly not the only one. For example, enterprise and text messaging, consumer social media, telephony and other data types are important content sources. This information should be included as part of the mix for understanding what really goes on in an organization and how employees interact with others.

## USING ARCHIVING PROACTIVELY

What decision makers need to realize is that their electronic archives are a wealth of information about the organization, and that archiving all relevant communications and content can help them to make better decisions about their business. The retention of electronic communications and other content – and the application of the right analytics to this content – can help decision makers to gain additional understanding about information flows that are simply not captured in other corporate systems. This combination of archived content plus analytics can help organizations to extract insight and intelligence about their business operations in ways that would otherwise not be possible.

While IT is focused primarily on costs and maintaining the critical systems on which an organization relies, demand for next-generation archiving will be coming from outside the IT organization: sales, marketing, customer experience, compliance and HR, since it is these organizations that will be leveraging the insights that can be gathered from archived data. Consequently, there needs to be a partnership among the key stakeholders inside and outside of IT to facilitate the deployment and use of next-generation archiving solutions. Every day that an organization is not archiving – defensively or proactively – is another day that they're missing data that could provide valuable insight to the business.

## USE CASES

There are various ways that organizations can use a combination of archived content and analytics to gain insight into their operations. For example:

- **Sales management**
  A customer relationship management (CRM) system is a useful tool to help salespeople manage the sales process. However, some have estimated that less than 10% of all relevant information is captured within CRM systems, since there is an enormous amount of data that cannot practically be entered by salespeople about customer and prospect interactions. This data includes the number of minutes between receiving an inquiry and responding to it, the tone of a customer's inquiry, and the various resources employed to make a sale. However, if information from an archive of electronic content could be captured and analyzed, it could answer a number of important questions about a company's sales process and success. For example:

  - How long does a salesperson take to respond to a customer's or prospect's inquiry or complaint and is there a difference in response time for a sales lead or inquiry versus a complaint?
  - Are sales teams working on the right opportunities? Are they prioritized correctly?
  - How long does it take for the customer or prospect to respond to the saleperson's response?
  - Is there a correlation between the length of time between the inquiry and the response to it and the likelihood of add-on sales for customers or initial sales for prospects?
  - Is there a correlation between the tone that salespeople use in email and other communication modes and the likelihood that a customer or prospect will buy?
  - Who are all of the other parties involved in a customer's or prospect's inquiries and sales' response to them? For example, do salespeople loop in others to help answer inquiries? Does this help in the sales process?

  In short, a combination of proper archiving and good analytics can dramatically bolster the efficacy of CRM systems and help organizations to understand their sales process more thoroughly. In addition, focusing analysis around an organization's successful engagements can help shine a light on techniques/best practices for overall team training.

- **Identifying potential data breaches**
  Data breaches are increasingly common and can result in major problems for any organization, potentially costing tens of millions of dollars to remediate. A next-generation archiving solution, in conjunction with a data loss prevention capability, can be useful in helping decision makers to identify anomalous information flows, such as employee accounts being used to send unusually large amounts of data or accounts accessing data sources they would not normally need to access.

- **Finding likely sources of employee fraud**
  A leading US-based law firm has discovered that employees who are treated poorly by their managers are more likely to send sensitive or confidential information to competitors, steal data or commit some type of financial fraud. One way to discover how managers are treating their employees is to analyze archived communications between managers and employees and between employees themselves. This could reveal that employees are being treated badly,

and could help senior managers to take proactive steps to stop managers from acting inappropriately.

- **Understanding and developing social relationships**
  As noted earlier, an archive of email and other communication contains information about how employees interact and with whom they communicate. Giving employees tools that will enable them to analyze these relationships can help them to find people within and outside of an organization who will help them to be more productive.

  For example, in the January 2016 survey of end users, we asked information workers how easy it would be to find a colleague who has an existing relationship with a client or prospect they would like to contact. Thirty-nine percent of the respondents indicated that it would be "not easy at all" to find the right individual, while only 10% responded it would be "easy" to do so. Moreover, employees are already seeking this kind of information, but through inefficient means: the survey found that the most common methods for finding others is asking peers in face-to-face conversations, searching the Web, looking in the organization's global directory, corporate intranet, or LinkedIn. Giving users a tool that would analyze archived communications and provide a "map" to a targeted individual could be very helpful.

- **Policy and legal compliance**
  Analysis of archived data can also be used to identify serious problems so that violations of the law, corporate policy or best practice can be addressed before they get out of hand. For example, a company's compliance officer could search for evidence of cheating, sexual harassment, lying to customers, illegal music downloads, distribution of pornography or any of several other activities that might result in a lawsuit, regulatory action, scandal or some other problem.

- **Understanding employee sentiment**
  Employee sentiment analysis can help decision makers surface key, recurring issues that are leading to low morale or high employee turnover (rumors of layoffs, for instance). If management can identify and understand the concerns that employees have, this might help managers to communicate more effectively with them and address problems before they can fester.

These use cases for an archiving solution are quite uncommon today for three reasons:

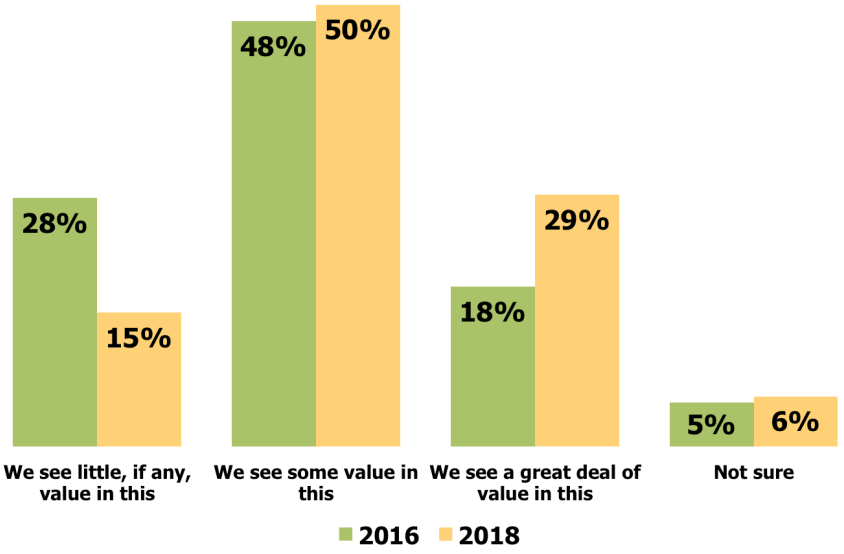
- Most decision makers view archiving as a primarily defensive capability and rarely consider the value of the intelligence buried in their archived data.

- Most archiving solutions will not support these types of applications.

- The decision makers most interested in these newer use cases are often outside the IT organization, and they may not be aware of the value that is in the data.

However, viewing archiving both defensively and proactively can yield important benefits, an idea that IT and non-IT decision makers alike will increasingly appreciate. As shown in Figure 3, a growing proportion of decision makers will see value in next-generation archiving over the next two years.

**Figure 3**
**IT's Perceived Value of Next-Generation Archiving**




*Source: Osterman Research, Inc.*

## MOST CURRENT ARCHIVING SOLUTIONS ARE NOT DESIGNED FOR ARCHIVING 2.0

There are three basic tiers of archiving solutions in use today, although there are relatively few next-generation solutions available:

- Legacy solutions that are typically deployed on-premises, offering reasonable performance for smaller data sets.
- Higher performance solutions – many of which are deployed in the cloud – that offer more robust performance because they were designed to handle much larger data sets.
- Next-generation solutions designed for both traditional archiving purposes and analytics of stored data.

### THE PROBLEM WITH CURRENT ARCHIVING SOLUTIONS

Legacy archiving solutions, and to a lesser extent some higher performance solutions, have a serious drawback in their design:

- **Data is stored in silos outside of corporate control**
  Data is typically stored in a large number of independent silos. Because of the proliferation of corporate applications, personally owned applications and cloud repositories, and IT's growing acceptance of the "Bring Your Own" trend, data storage is becoming highly fragmented and distributed. This makes data more and more difficult to access by legal teams, senior managers and others who must have access to it for formal purposes like eDiscovery or regulatory compliance. Ad hoc data inquiries, like those from individual employees in the normal course of doing their work, are also more difficult if an organization is storing a variety of data types. The result for end users is that they often do not look for their older data because of the time and difficulty involved in doing so.
- **Data is not easily connected across silos**
  The increasing distribution and fragmentation of corporate information leads to a

situation in which data cannot easily be connected across the various silos in which it is stored. This results in information access requiring visits to a number of individual data repositories one at a time – email, CRM, social media and several others – each with its own interface and learning curve. Moreover, data cannot be connected easily between silos, so a conversation starting in email and then shifting to instant messaging or some sort of collaborative tool cannot be viewed in context.

- **Storing data in silos opens the organization to significant risks**
  The ability to perform litigation holds, early case assessments and eDiscovery apply to all of the relevant data an organization might possess. For example, if an eDiscovery order demands production of all emails and other content from an individual, an organization is under obligation to produce all of its relevant data. However, if some of that data is stored in data repositories outside the control of IT, the organization runs the risk of being unable to satisfy the order. That can result in fines, sanctions or an adverse inference instruction in which a court can instruct a jury that failure to produce information can be considered evidence of culpability. While inaccessibility of data is often not an excuse, it is an increasingly common problem.

The bottom line is that many current archiving solutions simply were not designed with a focus on next-generation archiving. Our survey findings back this up: the March 2016 survey noted earlier found that only 10% of IT decision makers and influencers believe that their current archiving solution would support the use of analytics on their archived data.

## NEXT STEPS

Osterman Research recommends that decision makers take a four-step approach to archiving in general, and specifically with a focus on next-generation archiving:

- **Understand the traditional "defensive" role of archiving**
  First and foremost, decision makers must understand the value of archiving electronic content in its traditional "defensive" role. Archiving is enormously valuable as a means of retaining electronic content for long periods and ensuring that it remains unaltered and can be searched and produced on demand. For organizations that have been archiving for many years, this admonition sounds almost too obvious. However, many organizations still do not archive email and other electronic content despite archiving having been a best practice for many years. For example, an Osterman Research survey conducted in March 2016 found that only 69% of organizations have deployed a true archiving solution.

  Some decision makers believe that retaining "smoking guns" in email have the potential to create more harm than good; some are concerned about what they perceive as excessive storage requirements, while others believe that email and other forms of electronic communication are ephemeral and simply don't constitute records that the organization must retain. While regulations and legal precedent make it very clear that retaining electronic records is essential, some decision makers still don't buy into the notion of archiving. Often they don't see its value until there is a specific need or event (legal case, regulatory audit, etc.), but by then it's too late.

- **Implement the right data management approach**
  Next, decision makers must decide how all of their archived 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 is feasible, it presents some serious problems:

- There will be an enormous increase in storage requirements because the central archive will now contain a duplicate copy of archived content in other locations.

- At least some of the data in the central archive will be out of sync with data in the original archives as changes are made to the latter.

- Not all of the content in an organization needs to be archived, and so copying all data to centralized archives will store large quantities of superfluous data, unnecessarily adding to storage and storage management costs.

A superior approach is to implement a solution that will enable retention and analysis of content in place rather than moving it into a centralized archive. This reduces storage requirements, it makes data management much easier, and it allows easier integration of legacy data stores into the overall content repository.

- **Understand the benefits of next-generation archiving**
  The ability to mine information from archives of email, social media, files and other content sources is enormously valuable and can provide insight available nowhere else into an organization's operations. Analyzing this data can provide significant competitive advantage for any organization and will allow organizations to glean new insights from their available data. However, many of the benefits of next-generation archiving will be leveraged by stakeholders from outside of IT. Stakeholders need to be aware of both the value of the data they have and the potential applications for extracting insight from it. Decision makers need to be made aware of these benefits and authorize the effort necessary to realize them. That's not an easy task in some organizations, particularly in those that are not yet sold on the concept of traditional archiving.

- **Implement the right archiving solution**
  Finally, organizations should implement a robust content archiving capability that includes analytics capabilities. This will permit all relevant individuals to extract insight and intelligence from the archived content. A truly effective, next-generation archiving solution that delivers insight into an organization's operations must be able to retain information from a wide variety of data sources, including email, social networks, real time communications systems, CRM systems, blogs, point-of-sale systems, mobile devices and all other platforms and data types that contain or generate information about the business.

## SUMMARY

Archiving is an essential best practice for any organization, allowing it to defend itself in legal actions and during regulatory audits beyond helping users find the information they need for their work. The use of third-party archiving tools as supplements or replacements for native archiving capabilities is also a best practice in most situations.

However, archived data contains a wealth of information about information flows within the organization, how the organization *really* operates, the social relationships that exist between employees, and other valuable content. Applying analytics to archived data can deliver actionable results and competitive advantage in a way that no other solution can. It can enable organizations to utilize this rich data source in order to be proactive in identifying and addressing problems before they get out of hand, rather than purely for discovery and forensics after the fact. Every organization should seriously consider deploying a next-generation archiving capability that will enable it to extract insight and intelligence from its information archives.

## SPONSOR OF THIS WHITE PAPER

Micro Focus is a global software company with 40 years of experience in delivering and supporting enterprise software solutions that help customers innovate faster with lower risk. By applying proven expertise in software and security, we enable customers to utilize new technology solutions while maximizing the value of their investments in critical IT infrastructure and business applications. As a result, they can build, operate, and secure the IT systems that bring together existing business logic and applications with emerging technologies—in essence, bridging the old and the new—to meet their increasingly complex business demands.

**www.microfocus.com**
**@MicroFocus**
**info@microfocus.com**
**+1 866 464 9282**



135-000046-002