Why Email Archiving is a Critical Best Practice

An Osterman Research White Paper

Published October 2010

SPONSORED BY

Sonasoft

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

Executive Summary

THE SITUATION TODAY

Archiving – the set of technologies designed to support the practice of a) indexing email and other electronic content, b) placing it into archival storage and c) making it readily available via robust search tools – is a critical best practice, but one that many organizations have not yet implemented. For example, Osterman Research forecasts that by year-end 2010, only 47.9% of mid-sized and large organizations in North America will have deployed a *true* archiving system – the penetration of archiving among smaller organizations is even lower.

The fact that many small- to mid-sized businesses (SMBs) do not have a suitable archiving solution in place results in increased legal and regulatory risk, and it increases the overall cost of storage mangement.

WHAT PREVENTS ARCHIVING ADOPTION TODAY?

Osterman Research believes that there are three primary reasons that most organizations have not yet implemented an archiving system:

- Many decision makers believe that preserving content that could be used in a legal action or a regulatory audit is riskier than deleting that content. In a major survey conducted by Osterman Research in 2010, we found that 11% of decision makers believe the least risky approach for their organization is to delete all email content on a regular basis.

- Many believe that archiving content will create unmanageable storage problems and cost too much to manage, and so they delete content for fear of creating additional storage management difficulties.

- Many mistakenly believe that their nightly or other regular backups actually represent an archive of their important corporate content.

HOW DOES SONASOFT ADDRESS THESE ISSUES?

Sonasoft directly addresses these issues with a robust, enterprise-grade archiving solution that has been designed specifically to meet the requirements of SMBs:

- The solution archives all email content and so reduces the risk of non-compliance with legal, regulatory and other obligations to preserve critical business content.

- The solution will actually *reduce* storage management costs by eliminating redundant content and migrating content from email servers to lower cost archival storage.

- The Sonasoft solution is priced significantly lower than most of its competitors and is among the lowest priced archiving solutions on the market today.

ABOUT THIS WHITE PAPER

The goal of this white paper is four-fold:

- Discuss why archiving is a critical best practice for any organization.
- To dispel the mistaken notions that archiving is riskier than deleting all content and that archiving will create additional storage problems.
- To demonstrate that while regular backups are an important best practice, they are no substitute for archiving.
- To provide a brief overview of Sonasoft, the sponsor of this white paper, and their archiving solution.

Why Do You Need to Archive Email?

For virtually any company, there are five primary drivers for archiving, although the importance of these drivers will vary by the size of the organization, the industry in which it participates, the advice of internal and external legal counsel, and the locales in which it operates:

DRIVER #1: LEGAL CONSIDERATIONS

Email and other electronic content stores contain a growing proportion of business records that must be preserved for long periods of time. Further, this content is frequently requested during discovery proceedings because of the Federal Rules of Civil Procedure (FRCP) and state versions of the FRCP. As a result, it is critical that all relevant electronic content be made available for e-discovery purposes, in large part because of the frequency with which this information is requested during legal actions, as shown in the following table.

Corporate Actions That Have Occurred Related to Legal Problems

Activity	%
We have been ordered, as part of a legal action, to produce employee email	69%
We have referred back to our content archive or backup tapes to support our innocence in a legal case	57%
We have used archived content for pre-discovery purposes (i.e. to determine in advance whether or not to settle or fight a lawsuit)	49%
We have been ordered, as part of a legal action, to produce employee instant messages	14%
We have been ordered, as part of a legal action, to produce employee social networking content (e.g., employee Twitter or Facebook posts)	3%

Formally enacted in 1975, the FRCP governs court procedures for civil suits filed in the US federal courts. As a result of new amendments to the FRCP that went into effect in December 2006, the discovery of electronically stored information, including email messages, instant messages, word processing files, spreadsheets, presentations and other content, is now a mandatory point of discussion in civil cases. When subpoenaed for information, the responding party has a maximum of 30 days to respond according to Rule 34 of the FRCP.

The current version (2007) of the Rules requires the responding party to "[…] produce documents as they are kept in the ordinary course of business […]" Rule 34: 34(b)(2)(E)(i). This means that if the responding party uses data online and searches it electronically, they cannot supply that data as hard copy. The amendment also requires opposing parties to discuss e-discovery issues within 120 days of a lawsuit's filing.

When a hold on data is required, it is imperative that an organization immediately be able to begin preserving all relevant data, such as all email sent from senior managers to specific individuals or clients, word processing documents that may contain corporate policy statements, spreadsheets with auditors' opinions, and so on. An archiving system allows organizations to immediately place a hold on data when requested by a court or on the advice of legal counsel.

If an organization is not able to adequately place a hold on data when it is obligated to do so, it can suffer a variety of serious consequences, ranging from embarrassment to major legal sanctions or heavy fines. Litigants that fail to preserve electronic content properly are subject to a wide variety of consequences, including brand damage, additional costs for third-parties to review or search for data, court sanctions, directed verdicts or instructions to a jury that it can view a defendant's failure to produce data as evidence of culpability.

Further, an archiving system allows an organization to perform either formal or informal early case assessment activities. For example, if a terminated employee has threatened to sue his or her former employer in a wrongful termination action, senior managers can search the archive for information that will help them determine the potential liability they face. If this assessment of the potential lawsuit results in a determination that the company was indeed wrong in firing the employee (the aforementioned smoking gun), they can instruct legal counsel to pursue a quick legal settlement. If, on the other hand, the assessment results in the discovery of information that supports the appropriateness of the company's decision, that information can be used to convince the ex-employee to drop the case or it can help win the case if it goes to trial. In either case, the archiving system can help the organization to understand its position early on, either avoiding unnecessary legal fees or an adverse judgment, or reducing its costs by proving the sufficiency of its case.

DRIVER #2: REGULATORY COMPLIANCE

There are a large and growing number of regulatory obligations to preserve email. Some of the higher profile requirements are:

- *Health Insurance Portability and Accountability Act of 1996 (HIPAA)*
  All organizations operating in the healthcare field need to comply with HIPAA to ensure the safety of Protected Health Information. Organizations are required to protect the data from unauthorized users, as well as to retain for six years a broad range of documentation regarding their compliance.

  As part of the American Recovery and Reinvestment Act of 2009 (ARRA), the provisions of HIPAA have been significantly expanded. A key component of ARRA is the Health Information Technology for Economic and Clinical Health Act (HITECH). Now, business partners of entities already covered by HIPAA, such as pharmacies, healthcare providers and others, are required to comply with HIPAA provisions. This includes attorneys, accounting firms, external billing companies and others that do business with covered entities. While these business associates were accountable to the covered entities with which they did business under the old HIPAA, these associates are now liable for governmental penalties under the new law.

  Related to the point above is that penalties for HIPAA violations have been expanded dramatically. For example, if a covered entity or one of their business associates loses 500 or more patient records, they must notify HHS and a "prominent media outlet" to let them know what has occurred. Fines for violations can now reach as high as $1.5 million per calendar year.

- *Sarbanes-Oxley Act of 2002*
  The Sarbanes-Oxley Act of 2002 requires all public companies and their auditors to retain such relevant records as audit workpapers, memoranda, correspondence and electronic records – including email -- for a period of seven years. Company officers are obliged to report internal controls and procedures for financial reporting and auditors are required to test the internal control structures. Businesses have to ensure employees preserve information -- whether paper- or electronic-based -- that would be relevant to the company's financial reporting.

- *Securities and Exchange Commission Rules*
  Members of national securities exchanges, brokers and dealers are obliged to preserve all records for a minimum of six years, the first two years in an easily accessible place (SEC Rule 17a-4). The affected records are broad and encompass originals of communications generated and received by individuals within financial institutions, including inter-office memoranda and internal audit working papers. Also included are automated messages sent to all customers, which could include email blasts. The records may be "immediately produced or reproduced on 'micrographic media' [microfilm, microfiche or similar] or by means of 'electronic storage media'.

- *Financial Industry Regulatory Authority (FINRA)*
  FINRA is a non-governmental regulator formed in 2007 by the merger of various functions of the New York Stock Exchange and the National Association of Securities Dealers. FINRA manages a wide variety of rules that are imposed upon the more than 5,000 brokerage firms and nearly 675,000 registered representatives it oversees.

- *Model Requirements for the Management of Electronic Records (MoReq)*
  MoReq is a specification, originally developed in 2001, that defines the functional requirements for the manner in which electronic records are managed in an Electronic Records Management System. MoReq has been used widely in Europe and has been updated with MoReq2.

- A small sampling of the many other requirements for data retention include FINRA 3010, the Investment Advisors Act of 1940 (hedge funds), the Gramm-Leach-Bliley Act, IDA 29.7, FDA 21 CFR Part 11, OCC Advisory, the Financial Modernization Act 1999, Medicare Conditions of Participation, the Fair Labor Standards Act, the Americans with Disabilities Act, the Toxic Substances Control Act, the UK Companies Act, the UK Company Law Reform Bill - Electronic Communications, the UK Combined Code on Corporate Governance 2003, the UK Human Rights Act, Basel II, and the Markets in Financial Instruments Directive.

  The regulations above are but a very small sample of the regulations focused on data retention that impact archiving requirements and practices.

DRIVER #3: STORAGE MANAGEMENT

A company does not have to have billions of documents to experience significant email storage growth. The dual drivers of cheaper disk storage and the increased size of email messages, thanks to attachments such as images and videos, is fueling the electronic content storage explosion. Messaging storage is growing at roughly 30% annually, which means that a terabyte of storage today will swell to nearly 2.5 terabytes in just three years. Further, as shown in the following table, four of the top five problems in managing email systems are related to storage.

Problems Related to Managing Email and Other Electronic Systems
(% Responding a Serious or Very Serious Problem)

Problem	%
Increasing message size	55%
Increasing backup and restore times	51%
Using email as a knowledge store	41%
Lack of messaging-related disk space	37%
Mailboxes being overloaded	35%
Enforcing an email retention / deletion policy	35%
Protecting intellectual property	30%
Email continuity (making sure email runs 24/7)	30%
Lack of bandwidth	29%
Extracting content from .PST files (for Exchange users)	28%
Complying with government data retention statutes	25%
Poor server performance	24%
Managing policies handed down by senior mgmt.	23%
Keeping employees from being harassed via email	20%
Protecting the organization from wrongful dismissal claims	19%
IT difficulty meeting service-level agreements	19%

Archiving can be a very useful tool in reducing the volume of storage on email servers and in other electronic repositories, such as SharePoint or Quickr. One way to use archiving as a storage management tool is through the use of stubbing, in which email messages are replaced with "stubs" – roughly 10Kb links that point to content that has been migrated from users' mailboxes to the archive. When a user clicks on a stub, the message and attachment are retrieved from the archive and presented to the user as though the message were still in their mailbox. An alternative is to stub only attachments, leaving the message itself intact and replacing the attachment with a link. As with email stubbing, when a user clicks on the link, the attachment is retrieved from the archive. Another alternative is to migrate emails and attachments to the archive without the use of stubbing, allowing users to search for content directly from the archive.

Regardless of the particular method used, the advantages of using archiving to control storage growth include removing large amounts of content from "live" content stores like email servers or SharePoint servers and placing it on less expensive archival storage, as well as reducing the time required for both backups and restores.

DRIVER #4: END USER SELF SERVICE

Most IT staff members would wholeheartedly agree with the notion that users asking them to recover missing or deleted emails, files and other content is among the less pleasant aspects of their jobs. Aside from the difficulty associated with recovering this content, the time it requires takes away from other tasks that would allow IT to be more productive and contribute more to the company.

An appropriately designed archiving capability allows IT staff to put users in charge of recovering their own missing or deleted content, freeing IT from the burden of doing this for them. This can result in significant cost savings, as discussed later in this white paper, as well as recovery of IT time that otherwise would be spent on this important, but unproductive, task.

DRIVER #5: KNOWLEDGE MANAGEMENT / DATA MINING

As employees rely on email, collaboration tools and other content repositories as the primary tools they use to do work, it is important for companies to be able to extract business intelligence from the content that employees generate. Some archiving systems enable customers to quickly locate emails up to 15 years old and extract information, such as the identity of users' email correspondents or reports they generated. This could be useful, for example, when a new employee is required to trace back correspondence and other content between his or her predecessor and a customer. There are also sophisticated tools that can perform automated, large-scale retrieval and rigorous in-depth analysis of archived content, adding to the value of what archiving can offer a company.

BACKUP IS NOT ARCHIVING

It is important to understand that backup and archiving are two critical best practices, but they are not substitutes for one another:

- Backing up servers and other content sources is a best practice that is designed primarily to preserve a snapshot of data at a given point in time. The intent of a backup is to be able to quickly restore a server to normal operation in the event of a server hard disk crash, the application of a patch that does not go as planned, a power failure that causes a server to crash, or some other unexpected event.

- Archiving content, on the other hand, is designed to preserve information for purposes of activities like e-discovery, regulatory compliance or knowledge management.

It is also important to note that backups cannot preserve a complete record of data. For example, assuming that backups take place between 2:00am and 6:00am every morning, an email that is generated at 10:00am and then deleted by the sender and recipient at 3:00pm will never be included in a backup.

The fundamental difference between backing up and archiving, then, is that the former is designed as a tactical, short-term, *data*-focused solution that is important for restoring the proper operation of a server; while archiving is a more strategic, long-term, *information*-focused solution that is important for maintaining the integrity of all content generated or received by an organization. While both are best practices that any organization should follow, they are not interchangeable.

Key Issues to Consider When Selecting an Email Archiving Solution

There are a number of important issues to consider when evaluating and selecting an email archiving solution, among which are:

- **Ease of IT administration**
  IT departments are normally busy managing a large and growing number of systems, including email servers, email and Web security servers and/or appliances, encryption capabilities, and various types of other servers. Add to this the normal user handholding issues, problem resolution, product evaluation, configuration, policy management and other activities that IT must support.

  The result is that IT departments are very busy and, during a period of budget cutbacks, simply do not have the bandwidth to manage new capabilities. As a result, archiving systems must be very simple to manage and require as little extra IT staff time as possible.

- **Support for multiple content and export types**
  Archiving systems should support all of the content types that an organization needs today and will need in the future. This means that not only must email archived, but also content from Microsoft SharePoint repositories, user-generated desktop productivity applications, CRM systems, social media tools, instant messaging systems and other content sources. Further, a variety of export types must be supported to meet the requirements of senior management, internal and external

legal counsel, paralegals, opposing parties and others that may need access to data for legal, regulatory or other purposes.

- **Messaging system platform(s) supported**
  An archiving system should also support all of the messaging platforms that an organization operates. As the market leader in messaging, Microsoft Exchange is the platform that most organizations will want to support with an archiving solution, including Exchange 2010.

- **High availability**
  It is vital that any archiving system be operational as close to 100% of the time as possible. An Osterman Research study found that loss of archiving capability is even worse than email outages themselves. Loss of email that should be archived can carry with it a variety of quite serious consequences, including charges of spoliation of evidence and similarly damaging problems.

- **Use of multiple archive stores**
  The use of multiple archive stores is important to support both the performance of the archiving system, since an archiving system must keep up with the inflow of information presented to it; and disaster recovery.

- **Scalability**
  It is vital that an archiving platform be scalable in order to support the growing volume of content that it must support. For example, an organization of 500 users, each of whom generates 50 archivable messages of 150 kilobytes each on a typical workday, and that must retain content for seven years will generate an archive store of 45.5 million emails and 6.51 terabytes of content during that seven-year period. That doesn't count all of the other content sources that will feed into the archive that can boost these numbers by several times.

  As a result, an archiving platform must be highly scalable so that it can support the large quantity of content that it inevitably will need to house.

- **Robust search**
  One of the failings of many archiving systems is that they do not provide adequate search capabilities. This is particularly true in situations that require complex search queries, such as searching for any of 15 keywords in emails sent between any of 10 individuals between certain dates, while excluding other types of content. Search capabilities in an archiving platform must support very complex searches in support of various goals:

  - Rapid retrieval capabilities so that different scenarios can be tested.

  - Returning only relevant information so that the legal and related costs can be minimized.

  - Not returning too little information to avoid the appearance of withholding information.

- **Ease of use**
  Last – but certainly not least – is the requirement that an archiving system be easy to use. Because IT is already overworked in many organizations, non-technical people, such as paralegals and external legal counsel, should be able to create their own search queries without the aid of IT staff. If IT staff are required to set up complex Boolean search queries, for example, much of the value of the archiving system will be lost.

Moving to the Next Step in Email Archiving

UNDERSTAND THE CONSEQUENCES OF NOT ARCHIVING

Many decision makers, particularly those in less heavily regulated industries, may not believe that they need to archive email and other electronic content. However, there are a number of important reasons to archive that apply to virtually all organizations in all industries:

- An inability to produce all required content during e-discovery can result in sanctions directed against the party that cannot produce the necessary content. This might represent nothing more than incurring a judge's ire during legal proceedings, or it could result in significant sanctions. For example, in *Keithley v. Homestore, Inc.*[1], Keithley won on summary judgment, but still had to pay $283,000 in fees for its failure to produce required electronic evidence. In *Qualcomm, Inc. v. Broadcom Corporation*[2], Qualcomm initially prevailed, but it was later discovered that thousands of emails were not produced during the case. As a result, the Court awarded $8.5 million in attorney's fees and costs against Qualcomm. In some cases, judges have instructed juries that a party's inability to produce required content can be considered evidence of culpability.

- An inability to produce data on demand can result in the loss of reputation, brand damage or lost opportunities for future revenue. If, for example, a broker-dealer is sanctioned by the Securities and Exchange Commission for its failure to retain and produce data, the negative press that could result can harm the company's reputation with its shareholders and prospective customers.

- If an organization must go through an e-discovery exercise or a regulatory audit using only backup tapes or some other "quasi-archive", a substantial amount of time can be spent responding to the request for information. In addition to the direct cost associated with IT and legal staff extracting and poring through the data, there is also an opportunity cost associated with doing so. The result can be delays in other projects, postponement of infrastructure upgrades and the like, any of which can have consequences that may be difficult to quantify, but are nonetheless impactful.

- One of the primary functional benefits of an archiving system is its ability to act as a repository for older electronic content that might otherwise be stored on email

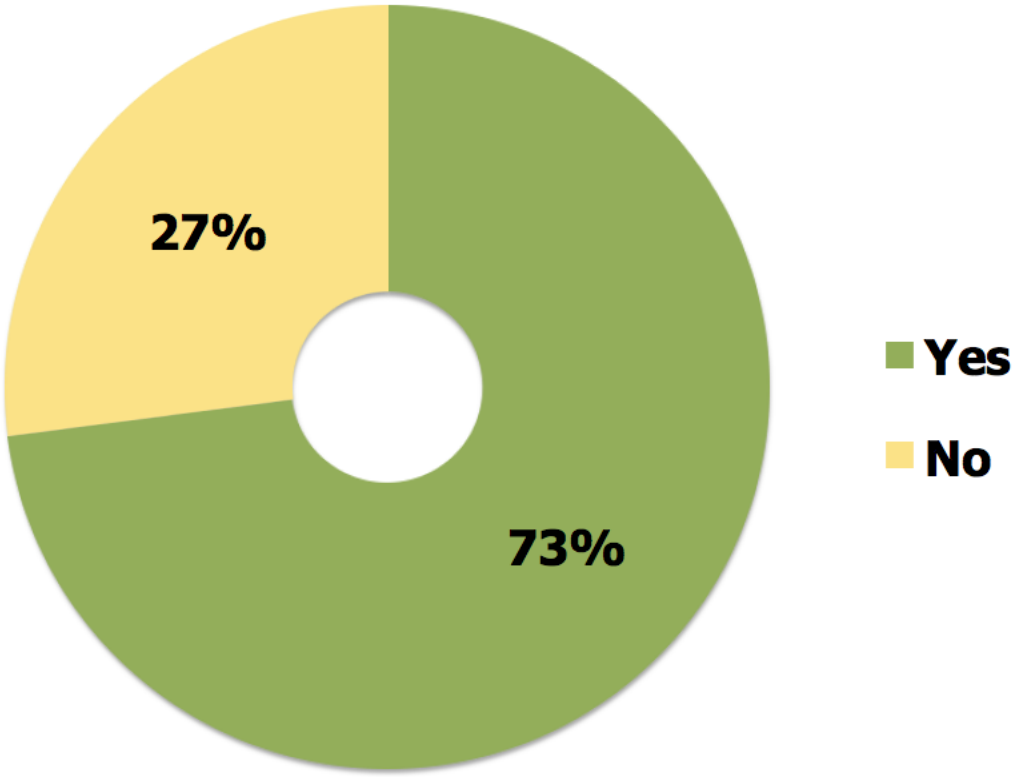
[1] Kevin Keithley v. The Home Store.com, Inc., 2008 U.S. Dist. LEXIS 61741 (August 12, 2008)
[2] No. 05-CV-1958-B(BLM), 2007 WL 2296441 (S.D. Cal. August 6, 2007)

servers or file servers. By migrating older content to an archive, email, file and other servers can operate more efficiently, backup windows for these servers are shortened, and restores can be accomplished much more quickly. For example, if 500 users are served by an email server that has gone down and the restoration process can be shortened by just one hour because an archiving system has permitted the data store to be smaller, that can result in a significantly lower productivity loss for those users.

ESTABLISH EMAIL AND OTHER CONTENT RETENTION POLICIES

The next step is to establish corporate policies focused on the retention of email and other electronic content. These need to be detailed so that specific types of content are preserved for the appropriate length of time as required by statute, legal precedent or corporate best practice. As shown in the following figure, many organizations have not yet established even email retention policies.

"Does your organization have an email retention policy in place?"



CALCULATE YOUR LONG-TERM ARCHIVING REQUIREMENTS

It is vital to consider long-term archiving requirements. For example, depending on the industry an organization serves, content archiving requirements can range from one year to indefinitely. That means that an organization could be expected to retain data for very long periods based on new regulatory requirements, legal precedents, industry best practice or simply the advice of legal counsel. In short, an organization should

expect its email and other electronic content stores to grow at a faster pace in the future, resulting in more content to archive.

DEPLOY A SOLUTION FOR THE LONG TERM

As noted above, content must sometimes be stored for very long periods, if not indefinitely. That means that an archiving solution must preserve content for many years and make it accessible to individuals that may have had no experience with deploying or using the archiving platform when it was originally installed. As a result, any archiving solution should be usable by employees for many years to come, meaning it must easy to use and update, as well as scalable.

About Sonasoft

Sonasoft was founded in 2002 for the purpose of providing software-based solutions that simplify and automate the process of replication, archiving, backup, recovery, centralize the management of multiple servers, and cost-effectively provide a disaster recovery strategy to protect valuable data. The result is higher availability of computer-based information, reduced risk of accidental or deliberate data loss, lower IT costs, and a sound disaster recovery strategy.

EMAIL ARCHIVING FOCUSED ON SMBs

Sonasoft's Email Archiver systematically records and saves information contained in email correspondence. Specifically, Sonasoft's email archiving technology supports:

- The ability to reduce IT administration costs by migrating all existing .PST files into a centralized archive and through the use of a Web-based interface that allows non-IT staff and others to access parts of the archive to which they have permission for access without the aid of IT.

- The ability to comply with the growing variety of compliance obligations in financial services, healthcare, energy and general business markets.

- The archival of content for all versions of Exchange from 2000 through 2010 (including the ability to migrate Exchange 5.5 .PST files), as well as public folder content.

- True single-instance storage that can reduce email storage requirements by up to 80%.

- The ability to manage multiple domains, making the solution useful for both hosted/SaaS/public cloud and private cloud deployments. UbiStor, for example, offers Sonasoft's archiving solution as either a SaaS or hardware-as-a-service solution.

- Robust litigation support capabilities, including e-discovery functions with an integrated query builder that can be used by non-IT staff, the ability to easily implement legal holds, and powerful case management capabilities.

- Robust reporting capabilities that can track trends in email usage, sources of data leaks, compliance with retention policies, storage compression ratios, access to the archive, etc.

Based in the heart of Silicon Valley – San Jose, California, the Sonasoft management team includes over 100+ years of combined experience managing the growth of high technology companies, as well as developing and marketing computer- and communications-related products. Sonasoft solutions are in use by a variety of leading organizations, including the University of Southern California, Kaiser Federal Bank, Tellurian Networks, Ubisoft, Moscone Center and many others.

SUMMARY

Sonasoft's Email Archiver offers a robust, enterprise-grade feature set that is designed to meet the regulatory and legal compliance requirements of small- to mid-sized businesses, as well as their needs for storage reduction, end-user self service for accessing archived content, and disaster recovery. The solution was designed for ease of use by non-technical staff members and is offered at an attractive price point. Plus, the solution can be deployed as a traditional on-premise system managed by internal IT staff, as a SaaS-based solution that can be managed by a third party, as a private cloud deployment, or as a combination of all three.