

[HOME](#) | [CURRENT ISSUE](#)

10 key considerations for email archiving

http://searchStorage.techtarget.com/magazineFeature/0,296894,sid5_gci1316576,00.html

by: [Stephen Foskett](#)

Issue: [Jun 2008](#)

Email archiving products vary in their features and technical structures. Here's how to select an archiving tool that's a good fit for your company's needs.

More and more companies are archiving their users' emails for business and legal reasons. If you haven't standardized on an archiving product, it can be a time-consuming process to find one that fits your company's needs; there are many choices available and each tool has unique features.

When examining an email archiving product, it's important to know how well it's suited to the specific requirements of the email system it's intended to protect. I've reviewed many of these products and compared their functionality to the requirements of dozens of companies. The following 10 questions will help you narrow down the available email archiving products to those that best serve your needs.

Not all of the following 10 questions will be important to every storage environment, but each one should be considered when making a product selection. You should decide whether or not a particular function is important in your environment. Not all email archiving implementations require legal-hold capability, for example. There can also be a spectrum of answers to each question, and not every environment needs the most extreme, feature-rich solution.

There are many considerations beyond the technical issues outlined here (see "*Email retention policy*," below). One of the primary deciding factors in any technology purchase is cost, which itself includes many variables. Vendor reputation, customer service and geographic support coverage may all influence product selection. While these factors aren't taken into account in this article, any one of them may have an impact and

must be carefully considered.

1. How complete is the archive?

Not all email archiving solutions capture every email, but that might not be desired. In some environments, only messages sent or received from the outside world need to be retained, so an email archive that uses a gateway approach would be acceptable. But many organizations require a more complete set of email messages, so the archive must interact with the mail server to ensure that all messages, both internal and external, are retained.

Even if an email archiving application captures inside and outside messages, some messages may still fall through the cracks. Archives that "sweep" through the mail system on a scheduled basis can miss messages that are sent, received and deleted between sweeps. Since every message has both a sender and a recipient, both of them would have to delete the message (and potentially empty their trash folder) to hide a message in this way, which is often called a "double delete" scenario. Organizations that are focused on compliance must ensure that their email archive captures every message.

2. Does it record what people do?

One step beyond a complete set of messages is an archive that maintains a record of user actions. Some systems are capable of recording whether a user opened, forwarded, flagged or filed an email message, a feature that has proven popular in product demonstrations.

However, "just because a message is marked as 'read' doesn't mean that a user really read it," says Matthew Ushijima, director of IT network operations at Empire Today in Chicago. "Outlook's preview pane can interfere in both positive and negative ways, making this [product feature] not the most reliable data source," he adds.

Capturing the actions users take regarding their email messages is a difficult technical problem. Traditional archiving products, which commonly use Exchange journaling, must sweep through the

mail system using MAPI to periodically examine each message to capture this so-called user-action meta data. MAPI sweeps consume valuable CPU and IO resources, so additional mail servers must be added to handle the load. An alternative approach to archiving, called log shipping, doesn't require these intensive sweeps, but is much less common. Consider whether this kind of user-action information is critical to your archiving needs.

3. Can the archive ingest an existing mail store or PST files?

Many organizations would like their email archive to include messages that existed before the archiving application was installed. These messages typically come from the mail system itself, which might include a decade or more of old mail, as well as from offline or user-created archives, like the PST files created by Microsoft's Outlook mail client. Many archiving programs are able to pull in these old messages, but some can't (see "*PST indigestion*," below).

Bringing in old messages from a mail server generally requires an intensive migration process using the MAPI protocol. This can take a few days, so the process is often performed over a weekend; large environments and those with email servers in multiple locations may find that it takes much longer.

Most email clients store personal archives on local disks, so these may be anywhere your users are, including laptops, desktops, network shares and portable drives. This makes importing archives tricky, as they must first be located and consolidated. Not every system can handle all formats, which can range from Outlook PST to Notes NSF, to Unix mbox and maildir files.

No matter where historic messages are imported from, the archive that contains them should be flagged as incomplete and potentially unreliable if ediscovery is a consideration. Both email servers and personal archives are almost certainly missing a great many messages. It's a trivial operation to change the content of most personal archives; modern email archive systems are far more tamper-proof.

4. Can the archive handle multiple email systems?

Not every email archiving application is capable of handling multiple email servers. If your environment features more than one email server, and especially if a variety of email systems are in use, this feature could prove critical. Generally speaking, archives that use a messaging gateway are far more flexible in heterogeneous environments than those that integrate more directly with the mail system.

This is especially common in organizations created as the result of corporate mergers, but some organizations find themselves in possession of heterogeneous mail systems for historic reasons. Whatever the cause, many email archive solutions don't support all of the various email servers, including Microsoft Exchange, IBM Lotus Notes/Domino, Unix mail and Apple's mail server.

5. What about non-message content?

Some email archiving applications focus only on messages, while others can also archive calendar items, tasks and contacts. A few also support other applications, including file systems, instant messages and database applications. Not every environment needs this type of archiving, but be sure to set expectations with management and your legal department about what is and isn't saved. While some archiving systems support content outside the email system, "email is the most critical," maintains Kelly Ferguson, senior product marketing manager for email archiving at EMC Corp. "Including file systems and SharePoint is nice, but email must get under control because it has the biggest risk due to message proliferation. Customers are starting with email, but have the expectation that the system can expand to other content types as need arises."

6. What about deduplication?

Deduplication is a hot topic and most email archiving vendors were early adopters of this capacity-saving technology. Email archivers that support deduplication will store only one copy of duplicate messages to conserve space and then link the messages in the archive. Some applications apply this only to entire messages, while others "crack" the message objects apart,

deduplicating attachments separately, which saves even more space.

7. Will the legal department be happy?

Although not all email archiving is performed to adhere to regulations, you must be prepared for a possible lawsuit that involves legal holds (this places a lock on certain emails) and ediscovery. Some archives produce exception logs and reports, and support extra-secure back-end storage to ensure that any content produced from them will satisfy the demands of litigation.

EMC's Ferguson points out that records of an archive system can be even more important than user meta data. "Who accessed the archive and what they looked at can be critical," says Ferguson. "This goes to the deletion policy as well; the system must keep track of every deletion that happens to prove that the archive is operating according to collection and retention policies." Some archiving applications can produce chain-of-custody reports for exported content, while others have security features such as encryption and SAS 70 security compliance audits (see "*Email supervision*," below).

If legal is interested in email archiving, they're probably looking for litigation-hold functionality. When a legal action seems imminent, they must instruct IT to hold a set of content in an immutable form in case legal discovery is required.

Some archiving applications include native litigation-hold features, but their granularity varies. Hold could apply to an individual object, a message, a folder, a user, a mailbox or an entire mail store, but not all systems can handle this variety. "Some can't place a litigation hold on individual items and need to hold an entire mailbox or message store to ensure that retention rules are stopped," claims Bill Tolson, director of legal and regulatory solutions marketing at Mimosa Systems Inc.

You should also determine if the system can handle multiple overlapping holds and change the scope of a hold without releasing it. Finally, the legal department might have different expectations about how to specify a hold; check with them to

see if they have any unique requirements.

Even if your archiving system of choice includes native legal-hold and search functionality, you may have to integrate it with a third-party tool. Not all archiving applications offer legal-hold and search features that are granular and flexible enough for real-world use, especially if your organization faces frequent legal action, so you might have to replace it with a more specialized legal-hold tool.

The process of declaring and releasing a legal hold can be complex. Holds normally cover a range of dates and systems, and the scope can change as discussions between the parties take place. Usually, a specialized third-party legal-hold program offers more functionality than one that's part of the archive application.

Another frequent objection to integrated hold-and-search features is the preference of the legal team. Ediscovery has become common in the last decade, so most attorneys have gone through the process a few times by now. It's likely their past experience with specialized litigation support software will lead them to request that solution instead of an unfamiliar one bundled with an email archiving system.

8. How does search work?

The search capabilities of archiving applications--which may be a critical feature for ediscovery--vary from product to product. Consider how your legal group conducts searches today; they may be very specific about what they're looking for or might want a more iterative process, such as nesting searches within searches until a set of messages is isolated. Try out the search function with the legal team to see if it works in ways they might want to use it.

"Our legal team is hooked on ediscovery," comments one administrator at a well-known business who asked to remain anonymous. "They are making many more queries and becoming better at asking more refined questions as they become familiar with the email archive's capabilities." This helps to protect the company and reduce ediscovery costs.

Consider the technicalities of the search function. Can it search across mailboxes or repositories? Requiring multiple searches might reduce the utility of the archive and content might be missed. Not all archiving systems can quickly and efficiently search a large data set. Try a query on a massive data set to judge the responsiveness of the archiver's search function.

9. Can the archive easily integrate with third-party tools?

In most cases, the email archiving system will become an essential part of your infrastructure, so consider how well it will integrate with other elements. Can the archive integrate with your user account management and access control system? What about your reporting, logging and audit tools? These features could become critical stumbling blocks as the product is rolled out across a large organization.

Note that the level of integration between archives and legal tools varies. Most archives can export a set of messages in a PST file for use by ediscovery tools, while others can directly tie in with the most popular tools with direct database access and APIs. The latter can be far more flexible and efficient, and organizations with frequent legal searches and favorite tools will benefit from this type of integration.

10. What will users think?

There's a spectrum of email client integration: some archives offer no integration and rely on a Web browser interface for archive access, while others use toolbars, executable extensions for certain clients or archive folders pushed from the mail server. Regardless of the technology used, consider the user's reaction (see "*What users want*," below). How will their interaction with their email client change once the archive is in place? If executable extensions are required to be installed on client machines, consider the impact of this rollout.

Think about the alternative email clients in use. Most organizations offer Web email clients, but some archiving systems don't integrate with those. Many users also access their mail using mobile

devices such as BlackBerry, Windows Mobile, Palm, iPhone and Symbian. However, most archiving systems have little or no mobile device integration beyond Web access, and these sites are sometimes poorly formatted or too graphically complex for mobile browsers. Off-line access is another key differentiator. If a user can access their archived messages while on a plane, they'll be far more likely to accept the system.

This article didn't cover the technical elements about product applicability to different storage environments, which are paramount considerations. Can, for example, the archiving application handle the size of your email system and the number of messages sent and received daily? Also consider whether the archiving product supports the operating systems and geographical layout of your email system. Not all email archiving solutions are able to scale equally.

Storage Magazine is a part of the [TechTarget](#) portfolio of enterprise IT-focused media. Copyright 2000 - 2008, TechTarget. All Rights Reserved.