Email Archiving: Do You Really Need To Archive Everything?

Data Storage Corporation
Excellence in Data Protection and Recovery

Message Logic is a business unit of
Data Storage Corporation. 212-564-4922
www.messagelogic.net or www.datastoragecorp.com

Email Archiving: Do You Really Need To Archive Everything?

"Half the money I spend on advertising is wasted," the great retailer John Wanamaker is credited with saying nearly 100 years ago. "The trouble is, I don't know which half."

IT managers and legal professionals could say the same thing about archiving email. Do you really need to archive everything?
Most regulations do not specify which messages need to be archived. Rather, they state what needs to be retrieved and, in most cases, how quickly they need to be retrieved. While the difference may seem subtle, the cost implications are huge.

Storage vendors add to the problem when they advise companies to archive everything in order to reduce legal liability. They correctly cite examples of significant fines and cases lost after needed email was wrongfully deleted.
On the other hand, if you save what you need to save and organize your archive for fast retrieval, you can reduce both storage and legal costs.

Storage vs. Legal Costs

The good news is that cost of magnetic storage has been declining at a rate of 45% per year since 1989. The cost of a terabyte of data, enough storage for 2000 scanned file cabinets, is expected to drop from $420 in 2005 to just $70 in 2008, according to Berghell Associates. Managed storage has already dropped to just 15 cents per gigabyte per month (Amazon's Simple Storage Service).

On the other hand, legal costs are escalating. When teams of attorneys and litigation support specialists review every email message for a case, the bills can exceed thousands of dollars per hour.

Therefore, while there are easy steps to reduce storage costs will save money, you may want to put the emphasis on more effective email retrieval. Even small improvements in retrieval accuracy may yield significant reductions in legal bills.
Fortunately, there are several meaningful steps you can take. The program involves identifying what must be kept, optimizing the retrieval of the most frequent requests, and determining what can be easily deleted. The steps are as follows:

1. Regulatory Requirements

The first obligation of any email retention schedule is to preserve email as required by government agencies for compliance review or for other regulatory and statutory reasons. Requirements vary by industry, geography, and company type. Your corporate counsel is probably aware of all of the requirements for your company. Here is a sample of mandated requirements:

- Sarbanes-Oxley requires accounting firms to keep records for seven years after an audit.
- HIPAA requires health care organizations to keep patient data for six years.
- Brokerage trading account records must be kept for six years after the account terminates.
- Medical records may need to be kept for two years after a patient's death.

Automated tools can identify documents that must be retained by sender, receiver, key word, and more. Optimized tools can go beyond the capabilities of search engines by using

email metadata. (Metadata includes the attributes such the sender, receivers, subject line, creation date, and routing details.) An optimized tool, such as Message Logic, can find messages sent to a particular company based on the domain found in the TO: and CC: fields. Most search engines would also find any message containing the company name in the text.

After the regulatory period has expired, counsel may advise you that records can be deleted.

2. Statutes of Limitation

The new Federal Rules of Civil Procedure protect companies when they delete email as part of "routine, good-faith operation." Unfortunately, the phrase "routine, good-faith operation" is not defined. The authoritative Advisory Committee on Civil Rules said that an entity would usually be protected if it took "reasonable steps to preserve the information after it knew or should have known the information was discoverable."

Clearly, the advice indicates that companies are not protected just because they follow a regular retention schedule. The length of the retention period must consider the relevant statutes of limitation and company contracts.

For example, Louis Testa, a truck driver for a fish wholesaler, unloaded a shipment at a New Hampshire Wal-Mart store. Testa slipped on some ice and snow that covered the dock ramp. He complained to a Wal-Mart employee on the dock at the time, but Testa did not take action for more than two years.

Wal-Mart routinely destroyed its records on the event in accordance with its two-year retention plan. However, the New Hampshire's statute of limitations on personal injury was three years. Testa filed suit after two years had passed. Wal-Mart could not produce evidence that it said included instructions sent to vendors informing them not to deliver merchandise that day. Wal-Mart lost the case. (*Testa v. Wal-Mart Stores*)

Wal-Mart had an obligation to keep messages as long as a suit could have been filed. If the retention policy had been three years for business-related documents or for ones where a complaint was made, there would not have been a problem.

Using the same logic, companies should identify the length of any contracts that might be contested in a court case. Emails may explain what was intended when the contract was written.

Deleting messages when an opposing party may have a copy could limit your defenses. Exact copies of incriminating email may be on desktop PCs, printed papers, BlackBerry handhelds, or the email server of an ISP. Courts have allowed plaintiffs to introduce printed copies of emails even though the defendant could not find an original in its system. (*Schwenn v. Anheuser-Busch*)

The result is that deleting messages may simply mean that you do not have access to evidence in a timely way and that any related messages that you could use for defense are not available.

It is a good idea to review the lengths of any key contracts and the statutes of limitation where you do business. It may be possible to delete messages after the time period expires. Consider exceptions listed in this document and check with legal counsel.

3. Litigation Hold

Ensure that you can override any automatic deletion policy. You must place a "litigation hold" on messages to prevent any evidence from being destroyed if litigation is "reasonably foreseeable." Some good indicators that a litigation hold is required are as follows:

- A formal complaint, subpoena, or notification of a lawsuit is received.
- Somebody threatens litigation, even verbally by saying, "I am going to sue."
- A regulatory or governmental body starts an investigation.
- An attorney or third-party investigator requests facts related to an incident or dispute.
- An incident takes place that results in injury.
- An employee makes a formal complaint to management, especially when related to personnel issues.

Take this requirement very seriously. Philip Morris was ordered to pay $2.75 million in sanctions when 11 managers deleted emails after litigation commenced. These managers, who had read the messages, were prohibited from testifying or disclosing the message content to the court. (*U.S. v. Philip Morris USA*)

Another court said that Samsung "willfully blinded itself" when it did not place a litigation hold on email. The court imposed $566,000 in sanctions and an adverse inference instruction. (*Mosaid Technologies, Inc. v. Samsung Electronics*)

When an incident occurs that may involve the courts, review the requirements with counsel. There are optimized tools available to identify relevant messages. You want to be able to do this by topic, sender (author), and receiver.

4. Unnecessary Messages

Once regulatory and statutory requirements, statutes of limitation, and litigation hold obligations are in the retention schedule, companies may try to reduce the number of emails stored or reviewed.

The next step is to create a process to identify common types of messages that probably would not be part of litigation and that your counsel says could be deleted as part of "routine, good-faith operation." It is critical to create a consistent plan with a defensible position about what messages are not needed. You may wish to consider (1) spam, (2) duplicate messages, (3) system notices, and (4) personal mail.

- Spam is probably the easiest to remove. Many spam filters can block such messages before they reach the email server or archiving system. Such systems are easy to deploy.
- Removing duplicate messages or attachments can save storage space. However, great care must be taken to remove only exact duplicates and to not alter messages that may be required for evidence. An alternative, which saves legal time, is to create a system that archives all messages, but allows them to be reviewed only once. In other words, if a copy of an email is marked as "responsive," all copies of the same message would also be automatically marked.
- Routine system notices, such as "the printer will be unavailable," may not need to be reviewed at all. If the company uses a standard format for system messages, they can be identified easily by most archiving systems. (Of course, it the availability of a printer is part of the case, you may regret deleting the message.)

- Companies may consider creating a short retention schedule or an automatic review for personal messages with no potential business impact. To identify personal mail, some companies ask employees to mark personal mail or to store it in a special folder. This is risky as it depends on employees to accurately decide what needs to be kept. It also can allow evidence to be destroyed if a rogue employee marks an incriminating message as personal.

There are not many systems that can automatically identify personal mail. Message Logic offers system that, when customized for a company, can often identify personal mail with an error rate is similar to that of many spam filters. In some cases, the customized personal mail detection system could reduce storage costs by 12%.

5. Shortening Retrieval

As storage prices are dropping, the most significant way to save cost may be to lessen the amount of legal time required to find relevant messages.

The popularity of search engines makes us accustomed to finding messages by typing in words or phrases. However, this is not the fastest or most complete way to find email messages.

- Google-like search engines are not effective for category searches. For example, it is not possible to ask for privacy leaks or offensive mail.
- Search engines cannot generalize. For example, while they can find email with a specific social security number, it is not possible to ask for emails with any social security number.
- It is also not possible to create a complete list of problem words in a category. All attempts will miss some words. In addition, in many cases, the individual words are innocuous. This leads to many false-positives.
- Searching through all the text in an archive each time a request is made takes too long.

Email retrieval can be faster if the system pre-processes messages using what it knows about common email searches. Most email searches (1) relate to sales, leaks, and employee matters, (2) have metadata with known formats, and (3) incorporate details about the business.

For example, searches for privacy violations can be made faster if a system identifies the presence of any social security number or other private content when the message is sent. Such messages can be "tagged" for quick retrieval. If it is necessary to investigate a privacy violation, optimized systems can quickly retrieve just the messages with tags instead of doing a text search.

Message Logic offers a system that pre-categorizes messages using 70 different tests, allows for customized tests, and fully indexes messages in real-time. In this way, messages can be retrieved dramatically faster when compared to a search engine.

Conclusion

Companies need to decide how much effort they want to put into managing retention. They can archive email forever, keep messages for the longest mandated retention period or statute-of-limitations time, or analyze each message and apply the appropriate time period.

While reducing storage cost is important, the most significant way to cut expenses may be to lessen the amount of legal effort required to retrieve relevant messages. Work with your legal counsel and be sure to keep the following in mind:

1. Preserve email as required by government agencies for compliance review or for other regulatory and statutory reasons.

2. Maintain messages for the time period of any statutes of limitation or contract period.

3. Ensure that you can override any policy if you must place email on "Litigation Hold."

4. Minimize storage and legal costs by minimizing the documents to be reviewed by legal teams: (1) spam, (2) duplicate messages, (3) system notices, and (4) personal mail.

5. Deploy systems that pre-categorize and pre-index messages to reduce legal costs.