Conducting Speedy Investigations that Involve Email

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

Conducting Speedy Investigations that Involve Email

"Timely detection': "rapid and current" disclosure, "conduct a reasonable investigation to promptly determine... ': "the most expedient time possible and without unreasonable delay': and "immediate and appropriate corrective action." These are just a few of the phrases used by legislation, regulation, and the U.S. Supreme Court to describe the time-sensitive
Requirements for effectively investigating a com- plaint, responding to a discovery request or governance and preparing compliance report.

Internal investigations usually require senior man- agreement to drop everything. Harassment and privacy cases may need immediate response. Sarbanes- Oxley responses have tight deadlines.

However, the need for fast response time conflicts with another major trend–the growing use of email as evidence. Most corporate investigations involve an analysis of email. One report in the National Law Journal states that at least 50% of the evidence presented in court cases is from email.

The reasons are clear. Email is the de facto journal of business activities. It is an uncensored, contemporaneous record of events and thoughts. Therefore, relevant messages can yield a gold mine of information for both sides in an investigation.

Unfortunately, finding relevant emails can be time consuming. Email files tend to be stored by date, not sender or topic. Therefore, even routine investigations may take days or weeks.

Companies may be penalized for delays or failure to produce timely information. Some recent high profile cases include the following:

- A jury awarded $800-million in punitive damages when Morgan Stanley repeatedly failed to produce emails in a timely manner. The judge stated that "efforts to hide its emails" were evidence of "guilt." (Coleman Holdings v. Morgan Stanley)
- A jury awarded $29.2 million in the largest single sex discrimination verdict in U.S. history after UBS Warburg could not produce copies of relevant emails. The jury was instructed to "infer that the [missing] evidence would have been unfavorable" to the defendant. *(Zubulake v. UBS Warburg)*
- The SEC imposed a fine of $10 million on Bank of America Securities, the brokerage arm of Bank of America, after they "repeatedly failed promptly to furnish" email and gave "misinformation."

Companies also have good reasons to respond quickly. Executives, CFOs, audit committees, corporate counsels, HR professionals, and compliance managers all have a stake and usually just a few days or weeks to act. Fortunately, systems can help and there are mistakes you can avoid to ensure speedy investigations and effective responses.

Retention Policy

The new Federal Rules of Civil Procedure (effective December 1, 2006) apply to any company that may find itself in federal court. For example, it applies to interstate contract disputes. The FRCP has serious implications regarding which records are retained for internal investigations.

FRCP Rule 37(f) protects companies from sanctions for deleting email as part of "routine, good-faith operation." The implication is that sanctions may be imposed if email is deleted in bad faith. Unfortunately, the phrase "routine, good-faith operation" is not defined. Certainly, any company with a policy of deleting all emails, or a 30, 60, or 90-day retention policy for the purpose of destroying smoking guns, ought to consider whether its policy would stand a court test of "good faith."

Even if a short retention policy passes a "good faith" test, it may not provide the protection such companies desire. 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 employer could not locate a record of these messages in its system.
(Schwenn v. Anheuser-Busch) In such a case, the employer cannot refute the evidence.

To create an effective retention policy for business email, companies should at least consider any man- dated requirements and the statute of limitations for any claims against the company.

Mandated requirements are numerous. 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 termination of the account. Medical records may need to be kept for two years after a patient's death. The last two requirements are tricky for IT as the retention period depends on an event, not just the calendar.

Statutes of limitations vary by state. One nationwide example is for the assessment or collection of federal taxes. The IRS sets the statute of limitations at three years after the filing of a return, unless there were misstatements, fraud, or evasion. For business email, companies need to decide how much effort they want to put into managing retention. One can keep all business email forever, set the retention period to the longest mandate or statute-of-limitations time period, or analyze each message and apply the appropriate period.

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

Few automatic systems exist to identify personal mail. Message Logic offers a personal mail detector that can be customized to automatically identify personal mail at many companies with near 99% accuracy. But, any automated system makes errors. As these errors are consistently applied, the Message Logic process may be considered "routine, good-faith operation."

This has not been determined by the courts and the company does not make legal representations about it. Message Logic reports disk storage can be reduced by about 12% of the total number of messages for a typical organization.

RECOMMENDATION: Follow the Federal Rules of Civil Procedure and create a retention policy that will stand the test of "routine, good-faith operation." Make sure that the policy considers the longest time period specified in relevant retention mandates and the statute of limitations. Consider an automated system for deleting personal mail to reduce storage costs.

TIMELY RESPONSE REQUIREMENTS (selected U.S.rules)-Chart 1

Name	Who It Concerns	What Is Required
Federal Rules of Civil Procedure (new rules effective Dec. 1,2006)	Any company that could be involved in litigation in a federal court. It includes all forms of interstate transactions.	Rule 26 requires disclosure of documents you may use to support claims or defenses without waiting for a discovery request. The parties must meet "as soon as practicable" to discuss "any issues relating to disclosure or discovery."Generally includes searches of emails sent and received by selected employees or by content.
Harassment Cases	Any company that could be involved with a hostile work environment claim (sexual harassment, ethnic harassment, religious harassment, etc.)	The U.S Supreme Court ruled employers may be held liable if the employer "fails to take immediate and appropriate corrective action."investigations of complaints often involve reading emails of those involved.
State Privacy and Identity Theft Laws (such aCA SB 1386)	Any company in one of 30 states with such legislation or any company that deals with residents in one of those states. (For example, any company with a customer in California.)	Various state laws require disclosure of data breaches, including email leaks, in the "most expedient time possible and without unreasonable delay."
Health Insurance Portability and Accountability Act (HIPAA)	Health care service providers and all health care entities, including insurance companies, government agencies, and, in some cases, benefits departments.	Enforcement rule 6.Section 160.412 says that the imposition of penalties can be precluded if the violation was not willful and the "violation has been timely corrected."

Email Retrieval

As soon as an incident, complaint, or discovery request takes place, the focus must be on responding quickly and completely. Penalties for delays can be significant. In one recent case, the U.S. District Court determined the appropriate fine for a late response to a discovery request was $50,000 per day. While the fine was eventually reduced, it was replaced by severe non-monetary sanctions. *(Serra Chevrolet v. General Motors)*

One common cause of delay is dependence on back-up systems instead of archive systems. Back-ups are optimized for business continuity, not email retrieval.

Back-ups are a recorded exact copy or "image" of an entire server at a specific moment in time. Because everything on the mail server is copied to the back- up, some organizations use them for a compliance record. The problem is that the process of retrieving a series of specific messages from back-ups is lengthy. It usually involves examining a series of back-ups taken from different days, weeks, or months. There are significant IT and billable legal labor costs to assemble the needed messages. In addition, to respond rapidly, IT professionals may be taken off of projects with little notice.

Archive systems eliminate these problems. Instead of containing a series of "snapshots," archives contain an indexed copy of each message. The index allows messages to be retrieved using a search engine. The best systems index entire messages, including attachments. They also allow searches on sub-sets of records, such as those from a specific sender or between a set of dates, to increase the speed and relevancy of results.

RECOMMENDATION: Invest in an archive sys- tem optimized for retrieval. While back-ups reduce up-front costs, the legal and IT cost of responding to the first request may significantly exceed the cost of the system.

Native Email Formats

In a landmark 2004 case, the U.S. District Court ruled that electronic documents must be produced "in native format" and "with their metadata intact." *(Williams v. Sprint)* Metadata includes attributes such as file owner, creation date, routing details, the sender, receivers, and subject line.

Therefore, it is important to be careful of steps used to reduce storage costs. The integrity of every message must be maintained. Common compression techniques, such as creating a ZIP file of messages, are effective, inexpensive and do not lock a company into a particular vendor. Proprietary compression techniques may yield incremental savings, but the cost savings are small.

TIMELY RESPONSE REQUIREMENTS (selected U.S. rules) — Chart 2

Name	Who It Concerns	What Is Required
Sarbanes-Oxley	Public companies, companies that may want to become public, and companies that may want to be acquired by a public company.	Section 409 requires reporting on a "rapid and current basis" of any material information. Typically involves a detailed review of emails regarding the largest sales and deals each quarter. Response must be "expedient, efficient, and thorough."
Foreign Corrupt Practices Act	Public companies, companies that may want to become public, and companies that may want to be acquired by a public company.	Department of Justice says systems must "provide management and the board of directors with timely and accurate information." Usually means that emails from sales people and executives are reviewed.
SEC Rule 17a-4	Financial Institutions and dealers.	Section (f)3 says that the party must "be ready at all times to provide, and immediately provide" records requested by the SEC.
Gramm-Leach-Bliley Act (GLB)	Certain Members of Broker-Dealer Firms.	Guidance on Section 501(b) states if there is an incident of unauthorized access to sensitive customer information, such as account information in unencrypted email, the institution must conduct an investigation to "promptly determine" whether the information will be misused.

When most people first try to find harassing or bullying email messages within a large body of messages they usually start by searching for dirty words and phrases. They quickly realize that additional types of words and phrases, such as ethnic slurs, need to be added. But, eventually, four problems emerge:

- They cannot think of all of the possible words and phrase combinations.
- They realize that some offensive words also have non-offensive meanings. The result is that the search yields many messages that are not actually harassment.
- They discover that as the list gets longer, the processing time to compare each message to the list gets longer.
- They discover that some messages that do not have any offensive words within them could be used as evidence of a hostile work environment.

Message Logic uses a more advanced, proprietary technique to find potentially harassing or bullying messages. These methods are primarily based on statistical language models. Message Logic assembled tens of thousands of messages from many companies and sorted them in terms of whether they contained potentially inappropriate content. We built statistical models of these messages to find which words and other elements are more commonly found in risky messages and which are more commonly found in messages that are not offensive. To analyze a new message, the Message Logic Archiver compares the message to the language models and performs a complex analysis to see if it is potentially harassing.

To demonstrate, Message Logic analyzed 500,000 messages sent and received by executives and professionals at Enron Corporation that were released by the U.S. government during their investigation.

The following message is similar to many messages that are common in harassment cases. This particular message contains a joke that employees could use as evidence to support a hostile work environment. Message Logic correctly identified this message. Other techniques would not have identified this message because it does not contain specific offensive words or phrases.

FROM: P***-
**..@ENRON
TO: R********/Corp/Enron@ENRON, K*******/HOU/ECT@ECT
DATE= 03/06/2001 TIME : 09:14:00
SUBJECT : Leaving
Early...

Three women all worked in the same office with the same female boss. Each day, they noticed the boss left work early. One day, the women decided that, when the boss left they would leave right behind her.

The brunette was thrilled to be home early. She did a little gardening, spent playtime with her son, and went to bed early. The redhead was elated to be able to get in a quick workout at the spa before meeting a dinner date.

The blonde was happy to get home early and surprise her husband, but when she got to her bedroom, she heard a muffled noise from inside. Slowly and quietly, she cracked open the door and was mortified to see her husband in bed with her boss! Gently, she closed the door and crept out of her house.

The next day, at their coffee break, the brunette and redhead planned to leave early again, and they asked the blond if she was going to go with them. "No way," the blonde exclaimed. "I almost got caught yesterday!"

Figure 1.One of many jokes circulated via Enron corporate e-mail.Note: Names were removed.

FROM : 8***..**
06/09/2000 02:28 PM
To: R******/HOU/ECT@ECT
Subject: Re: Kids

Either they have spare time or they are doing it in their sleep. I really don't want to think of anyone I know here working on having babies. I say that and yet I know Tracy is trying to get pg. She says she is tired of always having her legs in the air. I know she doesn't have any spare time.

Maybe she utilizes her time by doing two things at once. Like eating dinner and you know..... Or like, heck I don't know. My brain is mush. See ya. B

FROM: R*****
06/09/2000 12:54 PM
To: 8******/HOU/ECT@ECT
Subject: Re: Kids

You mean there are Enron employees with spare

time?? FROM: B*****
06/09/2000 10:57 AM
To: Robin R/HOU/ECT@ECT
Subject: Kids

Are you here yet?

There are thousands of kids here today. They are in every nook and cranny. Dang, I'll be glad to get out of here today. Are there thousands of kids on your floor too? We now know what Enron employees do in their spare time!! B

Figure 2. A conversation that does not contain dirty words, but might support a hostile work environment claim.

Above is an email discussion from the Enron email data. The two participants may not find the content to be offensive. It does not contain any dirty words or slurs.

However, this message could be offensive to many people. It could also provide supporting evidence in a case that does not involve the sender or recipient of the message. An attorney may discover the message in an email search. It could then be used as an example of the prevailing attitudes towards women, women who wish to become pregnant, or women who have children.

As with the other example, systems that depend upon lexicons or word lists would not detect this message. The Message Logic Archiver gave it a high ranking as potentially inappropriate mail that could be used to support a hostile work environment claim.

While examining products, be sure to look beyond the claims. Be especially skeptical of the products from companies that claim that they spent years working only on lexicons, word lists, and phrases. Ask for proof that the solution would catch these examples and others like them.

A powerful way to reduce storage costs is to remove duplicate messages. For example, it is possible to save one copy of a message sent to a distribution list. Pointers to the message are stored in the file of the other recipients. It is the exact same message in every way, nothing has been altered. However, de- duplicated messages must be exact duplicates. Even if the same message text *is* sent twice, the near duplicate may not be eliminated because the metadata is different.

RECOMMENDATION: Be extremely careful not to alter messages in any way. Reduce storage costs with archive products that use common compression techniques. Avoid proprietary compression that may yield only marginal reductions in storage and could lock the company into a particular vendor. Ensure that the de-duplication process only removes exact duplicates.

Using a Search Engine to Find Email

Most people search the Internet by typing a few words or phrases. However, this is not the fastest or a complete way to find messages in an email archive.

When searching emails, a system can take advantage of what it knows to improve 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. Optimized systems take advantage of this knowledge by preprocessing and categorizing messages. Searches can be made faster because the system already knows the relevant messages.

For example, messages with social security and credit card numbers can be tagged in real-time for potential privacy violations. Then, if it is necessary to investigate a privacy leak, optimized systems can act faster because they already know which messages contain these key risk factors.

The best retrieval systems allow custom tags in addition to the built-in tags. Example custom tags include competitor domain names and confidential project names. The best preprocessing systems take advantage of advanced techniques to tag messages that search terms cannot find. (See "What Language Technology Can Catch That Others Miss")

RECOMMENDATION: Many back-up and archive companies added search engines to enter the compliance market. While search engines demo well when the number of messages are limited, the retrieval process can be very time consuming when the number of messages gets large. Look for products that preprocess messages using advanced techniques and that are easily customized to substantially improve response time

Real-time Updates

Most investigations are based on past events assembled from archives or back-ups. But, if the problems continue, the company may be accused of failing to keep "rapid and current" or taking "immediate and appropriate" action.

Proper governance and fast action require current information. For example, for Sarbanes-Oxley, it may be important to know if a material event took place after the quarter ended. Systems based on messages from an archive or back-up snapshots cannot find new messages.

Preprocessing systems (see "Search Engine" on this page) that know when a critical event takes place can make a difference. The best systems act when a message is categorized. Actions should be customizable, and include alerts to management as well as advice emailed to the sender.

For example, alerts can be sent to notify HR when an offensive message is sent by employees who were previously warned about their behavior. Optionally, an automatic email can be sent to the sender to warn them of a potential problem. Alerts can be sent to management when mail *is* sent to a competitor or if a message containing the name of a confidential product is sent outside of the company.

RECOMMENDATION: For proper governance and for the ability to know if "corrective action" was effective, find systems that process messages as they are sent or received. Make sure that the list of events can be customized as well as the actions that can be taken by the system.

Don't Forget Internal Messages

Sixty percent of companies monitor external (incoming and outgoing) e-mail as a way to protect against intruders, leaks, and offensive content. However, only 27% monitor internal (employee to employee) messages where many violations are likely to take place. (American Management Association **/** ePolicy Institute 2004 survey)

"Management's failure to check internal e-mail is a potentially costly oversight. Off-the-cuff, casuale-mail conversations among employees are exactly the type of messages that tend to trigger lawsuits and arm litigators with damaging evidence," said Nancy Flynn, executive director of the ePolicy Institute, in a press release.

The reason is that many of the products in the market are designed for other tasks. For example, some companies with anti-spam firewall products use the same technology to monitor outbound mail. They are installed where the corporate network meets the Internet.

RECOMMENDATION: Avoid monitoring systems that install between the email server and the Internet, such as perimeter systems, firewalls, gateways or servers. These products may not process internal mail where most governance and employee matters take place.

Desktop Email Retrieval

Most back-up and archive systems are designed for use by the IT department. This makes sense because they are optimized for storing information in the system. The problem is that no matter how responsive IT wants to be, most IT departments are extremely busy. With only hours or days to respond, or when there is a tight quarterly fiscal deadline, the delays caused by the busy team can lead to fines, penalties, or sanctions.

The best email retrieval systems allow the people who need information to log in via a web browser. In such cases, the response can be in seconds.

RECOMMENDATION: Look for systems that allow those who need the information to get it at their desktops, without the need to wait for a busy IT department. Make sure that safeguards, such as an audit trail of all messages sent, are included to prevent abuse of the system.

Conclusion

Various laws and regulations mandate fast internal investigations for complaints, discovery requests, governance, or compliance reports. Executives, CFOs, audit committees, corporate counsels, HR professionals, and compliance managers may have just hours or days to get the emails they need.

Various products, such as Message Logic, are available as part of an overall program to monitor messages and to alert for potential problems. When being proactive, it is important to consider the following key factors:

1. Make sure that the company's email retention policies comply with the new Federal Rules for Civil Procedure. Every company, regardless of size, is affected.
2. Use an email archive system, not a back-up, for faster response.
3. Make sure the archive system does not alter email in any way or delete too many emails.
4. Avoid products based on search engines for retrieval. Preprocess messages by taking advantage of what is known about common requests, email formats, and the company.
5. Use products that process messages as they are sent and received.
6. Include internal emails to cover governance issues and employee matters.
7. Deploy systems that enable investigators to select and read email at their desks.