

THE STRATEGIC IMPORTANCE OF EMAIL ARCHIVE MANAGEMENT

An email archive strategy is imperative as email grows exponentially

By Kevin Craine, MBA

Every day thousands of email messages are created as part of a company's business operations. Organizations find that email clogs computer networks and demands increasing amounts of their employee's time. Email is growing exponentially, and as a result, vast quantities of email reside on network servers, PC hard drives, backup tapes, and other mass storage devices located in corporate information systems. The mounting volume of email, which in many cases must be archived, is the number one corporate storage hog.

Email also constitutes a "corporate memory" of how an organization conducts its business – evidence of corporate decisions and behavior. Email messages are testimony to an organization's functions, activities and transactions. Recent headlines draw attention to the role of email in the Microsoft antitrust trial and the multi-billion dollar settlement in the "fen-phen" diet drug case. Ask any corporate counsel what one worry keeps him or her awake at night and the likely culprit is the multitude of email messages stored throughout their enterprise. Email represents a growing and elusive risk in litigation, so understanding exactly what your corporate email says and exactly where it is stored is critical.

A great deal of "corporate knowledge" takes form in electronic mail. Nearly all organizations depend on email to conduct business, yet very few regard email as an important corporate asset. Mining the value of email requires a well-thought archive strategy. Email records can easily be destroyed or lost due to hardware failures and routine clean up, leaving behind only fragmented traces of the corporate knowledge contained within. Archiving and restoring email requires costly administration that is often inadequate, and takes a toll on employee productivity. And when a virus strikes the losses can be staggering, as was the impact of the infamous "I Love You" virus in May 2000.

It is now axiomatic that the World Wide Web will expand the dimensions of human communication and that the Internet will shape new markets. The emerging issue, however, is email.

Despite the expansive aspects of email, electronic messaging systems do very little to help users manage their vast and dynamic stores of corporate email.

Despite the expansive aspects of email communication, electronic messaging systems do very little to help users manage their vast and dynamic stores of corporate email. The first part of this paper attempts to demonstrate the escalating growth of email, the risks associated with this corporate information, and the knowledge value contained within ever-growing corporate email stores. The second part of this paper provides a matrix of desirable features and explores some fundamental issues to consider when selecting an email archive system. In the end, it is the hope of this author that you will be better prepared to make informed decisions and form beneficial strategic plans with regard to your corporate email environment.

The Growth of Email

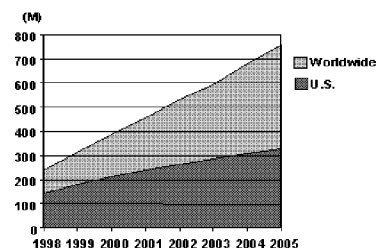
The growth of email use is phenomenal. Whatever watermark you use to measure the advance, this fact is clear: Email dominates the swelling tide of digital information – a tsunami of e-information that floods nearly every aspect of doing business and has begun to saturate corporate information storage systems.

Indeed, email has become so commonplace that it is hard to imagine getting along without it. This year, Americans will send and receive well over 6.8 trillion email messages – approximately 2.2 billion messages each day – compared with just 293 million pieces of first class mail.¹ Roughly two thirds of all American workers use email as part of their daily routine² and the average office worker will send and receive between 60 and 200 email messages each day.³ Microsoft, for example, processes more than 3 million email messages every day.⁴

The convenience of email has an infectious appeal, and people are spending more and more time sending, receiving and storing email. Workers polled this year by the Gartner Group, a leading e-market research firm, spent an average of 49 minutes a day on email, 30 to 35 percent more time than they did a year ago. Ferris Research estimates that management level workers spend four hours a day working with email. Indeed, many of Intel's 88,000 workers spend up to 2 ½ hours a day on email.⁵

Email is not just an American phenomenon, however. Traffic is growing even more rapidly around the world. The number of email addresses worldwide increased 84 percent in 1999 alone, with the segment located outside the U.S. leading the way with triple-digit growth.⁶ Today, over 225 million people worldwide can send and receive email,⁷ and predictions are that an astounding 750 million email addresses will exist around the globe by 2005.⁸

**U.S. and Worldwide
Email Mailboxes, 1998–2005**



The World Wide Web often grabs the headlines, but the fact is that there is 500 times more email created every year than the stock of Web pages (in terabytes).⁹ Although a portion of this email is sent and received by individuals, most of the traffic travels via corporate email servers.

Email Size on the Rise

While email use is growing exponentially, the size of email (i.e., bytes) is increasing as well. In the early 1990's, most email messages were between 1Kb and 3Kb in size, but the size of a typical email today (without an attachment) is around 5Kb. A decade ago file attachments were rare, but today over 20 percent of email comes with an attachment. Graphics, spreadsheets and Word documents frequently accompany email messages. As a result, email with attachments average around 100Kb currently.¹⁰ Analysts predict that the size of email and attachments will grow at a rate of 35 percent per year.¹¹

Because of this byte-sized trend, email has spawned a boom in the digital archiving business. Analysts estimate that the demand for digital storage grew by more than 1,800 percent between 1998 and 2003. Customers will pay dearly for this digital storage space. Spending on managed storage services, a mere \$11 million in 1999 is estimated to have grown to \$4.8 billion in 2003.¹² Revenues at EMC, the leading provider of digital storage, have increased well beyond forecasts – by over 50 percent annually.¹³ Increasing volumes of email are the cash cow in EMC storage farms.

To combat the ever-growing demand on server storage, system administrators often limit the allotted size of their users email store. Today, the median size of a message store allotted to a typical corporate user is 45Mb.¹⁴ If we assume that a typical user sends and receives 70 messages per day (at 5Kb) and 20 percent of those messages have attachments (at 100Kb), then our user generates a total of 1.75Mb of email traffic per day. If all of this data were kept in his message store, our user would exceed his allotted storage space in just over 25 days.

Users get around this limitation by storing email in multiple places (often multiple times). Email takes up residence in a hodge-podge of hard drives, disk drives, tape drives, laser disks, compact disks and network servers of all sizes and types. A midrange estimate of the amount of data currently stored on magnetic tape is 2.5 exabytes (an exabyte is 1 million terabytes) with another 2.5 exabytes stored on computer hard drives.¹⁵ Analysts estimate that 30 percent of stored email records and their attachments can be eliminated because they are redundant copies.¹⁶

Email Expense on the Rise

The quantity and size of email is growing beyond the ability of many organizations to effectively administer it. One study found that out of 926 organizations surveyed, 97 percent said that their email stores had grown faster than predicted. Nearly three-quarters had growth in excess of 50 percent, and a majority said that email backups consistently run into production periods. In 67 percent of the companies surveyed, restoring lost email for a single user takes over 24 hours to complete.¹⁷

The cost of email administration, therefore, is a growing concern. On average, an organization will spend nearly \$200 per user each year to locate and retrieve information from backup tapes. When considering lost user productivity, revenue loss and administration and management costs, companies spend more on this problem than for all technical support and help desk combined.¹⁸

Another factor that contributes to the escalating expense of email archive and administration is found in the nature and evolution of storage technology over the past 30 years. In the 1970's data archives were almost entirely transactional and interactive in nature with large mainframe systems processing airline reservations or health insurance claims. The 1980's ushered in the era of desktop computing and a shift in the nature of data being stored began to occur. While traditional transaction-based data still accounted for the lion's share of information in data stores, a growing percentage became static in nature (e.g., stationary vs. dynamic) as corporate email led the "knowledge worker" movement. In the 1990's this shift became more dramatic and today approximately 70 percent of data stored is static rather than interactive in nature.¹⁹

Due to this evolution in data file structure many of the current approaches to email archive management are unduly expensive because they are based on an outdated model. This cost can be a barrier to entry for many organizations. The "classic" approach of the 1970's and 1980's worked well for the type of interactive data prevalent 30 years ago, but implementing "big iron" may not be the best approach to archive static email messages...especially as volumes grow beyond expectations. As a result, some organizations find that the costs of an archive solution outweigh the risks of not having one in place. "We'd rather take our chances and pay the fines," laments one anonymous senior executive (see "risks" below). Any email archive solution you consider must be weighed against the expense and risks involved.

Experts estimate that for every dollar spent in storage for email archiving, another five dollars are spent to manage the system.²⁰ To assume that the cost of email archiving is equal to the cost of the associated storage systems is not entirely accurate. You must certainly analyze your volume in terms of terabytes, but you should also consider your archive volume in terms of number of objects that make up those terabytes. When combined, the volume in terabytes and the volume in email objects can dramatically impact your expenses, as a large database and search engine infrastructure must be implemented to enable the accurate retrieval of such objects. Traditional systems are focused on reducing the expense of storage hardware, but do nothing to reduce the expense of infrastructure and administration.

In short, email messages are growing in quantity and contain increasing amounts of information. The time that employees spend using email is on the rise and more time and money is spent than ever before in attempts to manage the ever-growing tide of email. Traditional approaches to data archive may be outdated and cost-prohibitive to implement.

The Risk of Email

Email is more than just bits and bytes, however. Email is a cornerstone of corporate business and has quickly replaced written documents, faxes, and even the telephone as the primary communication conduit for most corporations. As computer technology has advanced and people have become accustomed to using email, the importance and complexity of email has evolved from a few simple text messages into a multitude of mission-critical documents. Negotiations, bids, proposals, contracts, legal agreements, regulatory forms, and a host of other vital correspondence now find form in email.

But with this progression comes increased corporate risk. Email is essentially a "corporate memory" of how an organization has conducted itself. A potent and powerful factor during litigation, courts in the U.S., as well as many other countries, have ruled that email relating to potential litigation must be preserved as evidence and is subject to disclosure. According to Richard Lazar of Fios Inc, an electronic discovery services company, "email can often show - with a high degree of reliability - who said and knew what." Lazar, a former litigator and now President and CEO of Fios, points out that "email often shows who else was involved in activity. It is a litigator's primary weapon in the search to uncover a needle in a haystack."²¹

A potent and powerful factor during litigation, courts have ruled that email relating to a particular criminal case must be preserved as evidence and is subject to disclosure.

As email becomes the default method of business communication the ability to archive, search and understand vast stores of email with effectiveness and ease is imperative in order to mitigate corporate risk during litigation. While email is often used for formal correspondence, the informal nature of email can bring additional risk. Email reaches beyond traditional business communication and bonds people in a web of informal banter that is potentially dangerous. Employees tend to be less formal when using email even when discussing formal business and often unknowingly treat email as an unrecorded telephone conversation.

The essential questions are: What does your email say about your company and the conduct of your employees? Would you be comfortable having your email exposed in a court of law, tantamount to a retroactive wiretap on your communications?

Email is the "Smoking Gun"

For trial lawyers email often represents the "smoking gun" that will win their case or force a settlement. Attorneys aggressively lobby courts for access to their opponents email stores. This in itself can be a significant corporate risk. One executive bemoans: "We had to produce all the email pertaining to the matter from the past 10 years. This was a tremendous expense. Some of the information was on tape backup, but was very difficult and time consuming to select, filter and reproduce. We also had to go to every user to get local email stores from laptops, desktops and even some employees' home systems. It was a nightmare and cost the company millions."²² Judges have noted the costs and effort associated with retrieving email records, but have characterized this burden as "one of the risks taken on by companies which have made the decision to avail themselves of computer technology."

Consider the case of a female employee who brought a lawsuit against software giant Oracle. The woman had a romantic relationship with CEO Larry Ellison, but was eventually fired for an unrelated cause. In retribution, the woman sent an email message from the account of another Oracle executive

saying, "I have terminated [the woman] per your request." Based on the email, the employee filed a wrongful discharge and sexual harassment suit against Ellison. But the district attorney found the smoking gun in the Oracle email audit logs, and eventually proved that the woman had falsely created the email message. She was prosecuted for perjury and falsifying evidence.²³

Perhaps the most popular example of damning email in litigation is the U.S. Department of Justice's antitrust trial against Microsoft. The content of email traffic between Bill Gates and other Microsoft executives was startling and captured attention in many corporate boardrooms. This message rang clear: If a damaging email exists, someone will find it, and it will most certainly resurface to haunt you.

To mitigate the potential risks associated with discovery, some companies have a policy to intentionally delete email messages (after 30 days, for example). This type of approach can backfire, however, as was the case in the "fen-phen" diet drug suit. One of the manufacturers was severely sanctioned for overwriting tape backups for several months after litigation had begun. The court instructed the jury to infer that the deleted messages contained evidence that was potentially damaging, and that the company had either negligently or intentionally destroyed the evidence.

Another risk that is often overlooked is the fact that a great deal of personal email is transmitted via corporate messaging systems. This can leave a company open to both litigation and theft. The free flow of email means the backdoor is potentially always open. One example is a recent case that involved a contract employee at Gillette Company who was caught using email to sell stolen plans for the company's new Mach-3 razor.

68% of companies surveyed characterized email abuse as widespread, with losses estimated at \$3.7 million per company per year.

- Lee Bruno, "How Safe Are Your Business Secrets?"

"30% of the email going through our servers was not work-related."

- Jeff LePage, director of MIS at American Fast Freight Inc.

In short, the risks associated with email are significant in terms of litigation, exposure of corporate and employee behavior, and theft of proprietary information. In order to mitigate these risks, companies must understand where all of their email records are located, what they say and who is saying it.

The Knowledge Value of Email

A great deal of "corporate knowledge" takes residence in the form of email. About 35 percent of the business information used in day-to-day business is stored in email messaging systems.²⁴

Firms are beginning to recognize email as a corporate asset, but for the most part they have been powerless to gain control or do anything constructive with this intellectual property.

Mining and managing the information contained in email is a challenge that is more demanding than simply moving bits and bytes to tape. The key to finding the "knowledge value" of corporate email lies in the ability to access, understand and use these archived messages in beneficial and collaborative ways. Email records need to be quickly and easily captured and archived so that they become part of the available corporate information repository.

Hidden Knowledge Assets

Important email information is often hidden from the rest of the organization. For the most part, individual users decide which email messages will be retained or deleted. This is especially true with older email

