

## **Leveraging the E-archive**

September 22, 2008 *By Charles Weeden*

Over the past five years, most institutions regulated by the Securities and Exchange Commission have implemented compliance archival technologies. However, this is likely only the first step of a three-phase evolution that, in the end, will ensure that financial institutions are leaders in managing and leveraging unstructured intellectual property, commonly referred to as electronically stored information, or ESI.

Though Wall Street historically has led other industries in the adoption of new technologies, these adoptions were the product of internal, profit-driven decisions. The adoption of archival technologies, however, is the result of regulation and while financial institutions resisted implementing these technologies, ultimately the beneficiary will be the institutions' own bottom lines.

During 2001 and 2002, the SEC staff debated whether e-mail should be considered customer communications and thus subject to Rule 17a-4 of the Securities and Exchange Act of 1934. Forced by revelations that emerged from e-mails during the dot-com crash, the SEC decided that e-mail and other e-messaging records--instant messages (IM), Bloomberg and Reuters messages, etc.--are subject to regulatory retention. Phase one, or the deployment of e-mail archive technologies, has been completed at many firms, and most IT and compliance managers mistakenly see this as the final step of the project.

Typically, current archival technologies forward a copy of every e-mail from the institutional e-mail system--we'll use Microsoft Exchange for the purposes of this article--into the archive server. Thus, one copy of the e-mail is retained on the Exchange server and one is retained in the archive. Though this architecture may seem needlessly duplicative, savvy IT administrators realize that the archive offers solutions to many of their pressing information management problems.

Most users do not spend time managing and, more importantly, deleting e-mail. This causes the Exchange servers to become bloated repositories of old, unnecessary communications. Operationally, the size negatively affects the overall efficiency of the e-mail system by demanding time and resources to backup e-mail records that are no longer useful.

Over the years, Exchange administrators have tried many strategies to reduce the size of the information store, including imposing mailbox limits or forcing users to locally save e-mail records on their workstations. These approaches did little to stem the tide, and the size of the e-mail store on the Exchange server continued to grow. Worse, senior management did not feel they should be subject to these limitations. In fact, in most firms the size of executives' mailboxes are by far the largest, and they sometimes claim to have every e-mail they have ever sent or received, without realizing the cost incurred in backing up these files each night.

### **Sophisticated Searches**

In phase two of the e-archiving evolution, Exchange administrators will take advantage of the compliance copy of e-mails. We've seen many institutions delete all e-mail on the Exchange server after, perhaps, one year and require users to search for older e-mails in the corporate archive. Users then discover that the search technologies on the archive system are far more sophisticated than those on Exchange or

Outlook, so many choose to use the archive for all search functions. The archive also incorporates users' other e-messaging, such as IM and Bloomberg messages, allowing them to search all their e-messaging from the past seven years. Phase two, therefore, leverages the archive by providing users with both a better search tool and a comprehensive repository, allowing the administrator to dramatically reduce the amount of e-mail stored on Exchange. Exchange servers can then be configured to do what they were designed to do--send and receive e-mail-not serve as e-mail warehouses.

Industry pundits have long held that 70 percent of an institution's "knowledge" is retained in e-mails. Users are also aware of this, but firms have found it difficult to leverage this knowledge on a corporate level. However, several catalysts are present in the financial industry to help firms transform and mine this unstructured data: First, most content and communications are now converted to an electronic format; second, employees accept that this content is "owned" by the institution. In addition, thanks to Rule 17a-4 and other regulations, virtually all unstructured content is retained. This sets the stage for phase three of the evolution--technologies to make an institution's "knowledge" available to the institution.

### **Corporate Knowledge Base**

During phase three, a firm brings users' search and information management tools to the entire organization. Users may now register e-mail records as available to their department or the general corporate knowledge base. Users also can submit their e-mails to corporate blogs for a broader discussion. Once this phase is implemented, a user's search may extend to the entire corporate knowledge base, including department wikis and blogs. E-mails are reduced and corporate knowledge leveraged.

In a phase-three environment, some estimate that 30 percent of an institution's e-mail will transition into wiki or library searches. Though traditional enterprise software and corporate manuals will still dictate standard policies and procedures, wikis and blogs--created by users to deal with specific situations or exceptions--will be the predominate tool for communicating non-traditional information across an organization.

In our experience working with financial firms, we've seen the beginnings of phase two. Many firms are successfully weaning their users off of their Outlook mailboxes and teaching them to use the search and organizational tools for the archival copy. Users have become comfortable using more sophisticated search tools including such options as wildcards, proximity, and even Boolean logic. And the next generation of search technologies, including context and semantic searching, is just now beginning to be available for unstructured repositories.

With thanks (albeit grudgingly) to the SEC, we believe that U.S. financial institutions will be the first to adopt these technologies and implement comprehensive programs to empower and leverage unstructured content. While European Union institutions argue about whether corporate e-mail records are subject to individual privacy concerns, and other countries debate whether unstructured content is a part of corporate records, U.S. firms will leverage their lead, mine the buried 70 percent of corporate knowledge, embrace wikis, blogs and other grass-roots knowledge bases and become the most globally efficient users of the most important form of capital--corporate intellectual capital.

*Charles Weeden is managing partner of 17a-4, a New York-based e-messaging compliance consultancy.*