

# Alchemy MailStore for Exchange

*Automating Email Capture for Compliance and Records Management*

**October 2003**

**A White Paper by Greg Council**

Product Manager  
Information Management Research Inc.



“As e-mail gains momentum as the communication tools of choice in companies worldwide, CIOs and IT organizations must be knowledgeable of the regulations and aware of the solutions available to help them meet these requirements.”

Today's businesses are wrestling with a major e-mail issue. Once considered only a convenience application for communication, email is now the life-blood of most enterprises and is considered mission-critical. The result is a tremendous reliance of email as an official medium of communication.

But this use has led to the realization that email messages and their attachments are now subject to legal scrutiny and the many government requirements that treat them as an official record.

While compliance mandates appear to be concentrated on specific industries such as healthcare and financial services, public companies across all industries are finding themselves subject to records retention regulations.

“As e-mail gains momentum as the communication tools of choice in companies worldwide, CIOs and IT organizations must be knowledgeable of the regulations and aware of the solutions available to help them meet these requirements.” – Charles Brett, senior program director with META Group.

MailStore for Exchange is devoted to tackling the email archive and records management issue that is becoming prevalent in just about every corporation, whether or not they are regulated.

## Overview

The MailStore solution has three distinct modes of functionality, each designed to allow maximum flexibility for solving the unique problems encountered by each organization:

- Server-based archiving and message records management. The purpose is to automatically copy a “snapshot” of messages sent and received through a mail server and store them into a central, secure archive repository. Automation and security are keys – no one can tamper with the process, an absolute must for meeting the burden of proof in legal situations.
- Client-based access. The purpose is to give end users the tools to search and view the message archive without IT intervention.



- Client-based archiving. The purpose is to enable Outlook end users to archive the contents of their PST files into a central, secure repository, often one that contains other types of documents.

### **Server-based Archiving Process**

The MailStore for Exchange philosophy is “maximum results with minimum impact on the mail server”. In order to deal effectively with mail system peak times and minimize the impact on the mail server, the archiving process has been split into three services.

1. Capture messages from email server. The Tracking Service is a non-invasive Exchange Server-resident application whose sole purpose is to listen for any incoming or outgoing messages and create a copy for automated processing by the Archive Service. Exchange administrators who want a totally hands-off installation can also implement MailStore’s Exchange Journaling archive option, the preferred method for Exchange 2000 and 2003 systems.
2. Process messages into the archive repository – The Archive Service provides rules creation, management, and processing. Each message that is copied is processed according to rules the administrator creates and then, based upon those rules, the message is either archived or the temporary copy is deleted message copy.
3. Manage Message Access (Alchemy Server) – The archive service accesses the archive(s) via a repository service called the Alchemy Server and makes the appropriate entries to the data container and indexes. Alchemy Server also manages user access controls and repository publishing.

Each time Exchange reports a transaction, the Tracking Service (or Journaling) creates a copy of the transaction record. This REAL TIME mode of capture insures the authenticity of each email record without risk of an end-user deleting or modifying it before it can be archived. The original message is unaffected and delivered to the recipient(s) with no adverse affects to performance.



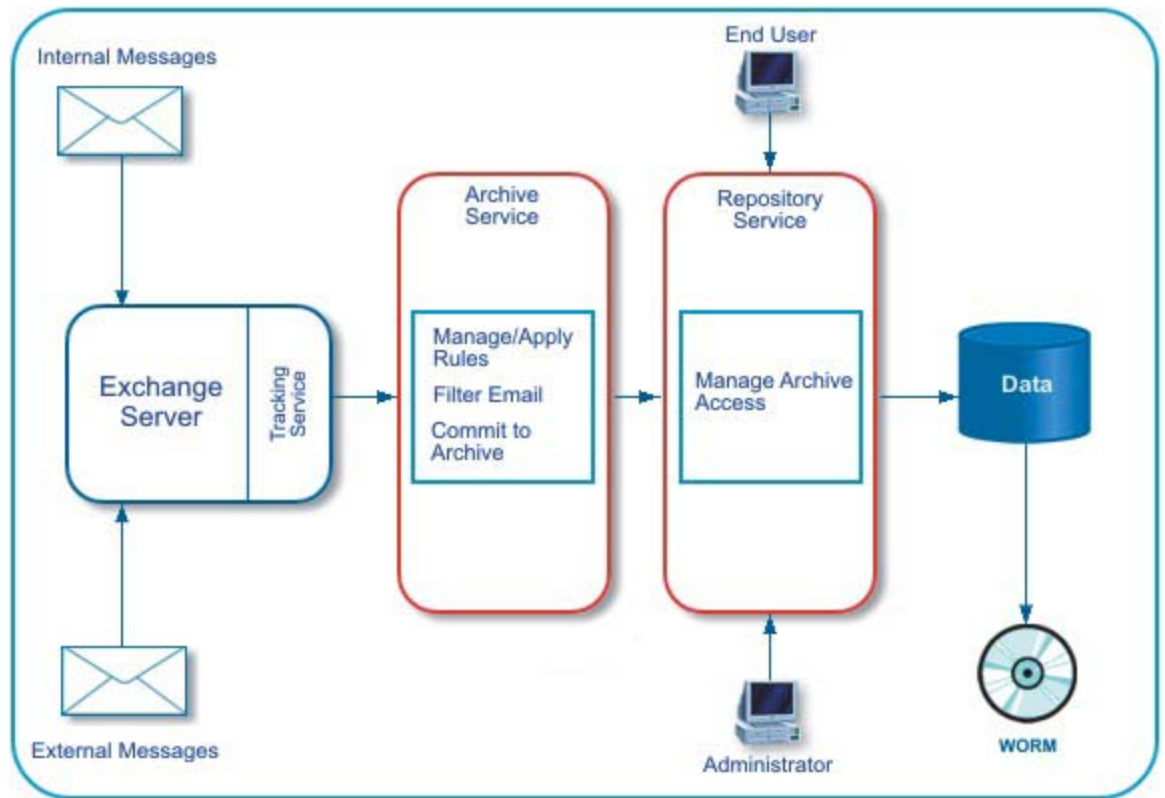
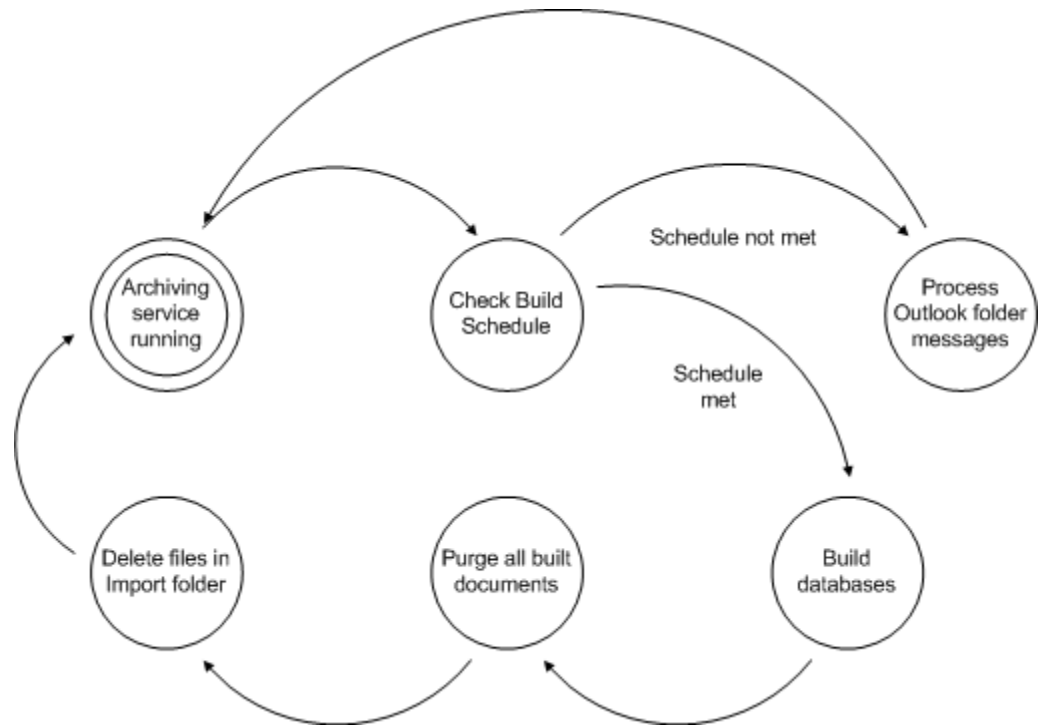


Figure 1. Email Archive Flow

The Archiving Service moves the temporary copies into a secured system folder typically located on another server. The administrator can configure the timing of when the Archive Service processes the messages – the default is to process each message immediately as it comes through the system.

The following diagram/process illustrates how messages move into the archive.



1. Administrator configures the Build (Index) Schedule.
2. The Archive Service checks schedule against current system time once per minute.
3. If schedule conditions are met, the build process is initiated.
4. Each archive repository specified in the databases list is built according to the settings currently specified by the administrator.
5. Once the database is built, all temporary copies are deleted.
6. The archive service resumes the processing of new messages.

## Message Metadata Processing

During the Archive Service process, information is extracted from the email to provide data for a number of processes:

- ❑ Classification and Search Metadata – by default, all messages have their primary header data extracted and populated into a message metadata index. The fields used for this index are the “To”, “From”, “Date”, and “Subject” header data. Since this data is passed to the repository in real time, within several seconds of capture each message is searchable from the repository using these fields.
- ❑ Rules Metadata – if processing rules have been created, the Archive Service will examine the extracted data to determine how the message is to be handled.

## The Rules Engine

MailStore for Exchange has a very powerful, flexible, yet easy-to-use rules engine. It is capable of parsing any alphanumeric pattern in the message header data extracted, and then triggering an action or set of actions based on pre-determined rules. Standard MailStore for Exchange rules include ‘archive all’, ‘archive only external messages’, and ‘archive selected mailboxes’. Rules can also be set to exclude messages that are not deemed necessary for the archive. Examples include news-lists, spam, personal email sources or virtually any domain or email address the administrator wishes to exclude. Filters can apply to all or some archived mailboxes.

Specifically, rules can be created to perform any of the following actions:

- ❑ Archive/filter message by domain or email address
- ❑ Archive/filter message by exact word or phrase in subject
- ❑ Archive qualifying message to specific message repository
- ❑ Archive/filter specific mailbox



Each of the above can be combined to create a comprehensive set of rules to enforce the most demanding corporate email archive policies.

## **Message and Attachment Content**

As part of the archive process, the content of each message and attachment is extracted and examined.

If a message contains an attachment or attachments, each is stored as a compound document format where the message is the parent and the attachments are the children. This format enables several benefits for the end user who searches the message archive. First, all text-based attachments are available for the full text index. Second, when one attachment is retrieved via full-text search, the user can also view the message and associated attachments. Finally, browsing the archive for attachments is very easy as each attachment is contextually linked and visible.

The content of every message is extracted and added to the full-text index according to the Build schedule mentioned earlier, as is the content of any attachments. The message header data (To, From, Date Sent and Subject Line) is full-text indexed as well, providing very powerful search capabilities not limited to a fielded query. Once added, the message and attachment content is available to the built-in search engine of the administrator application. The header data is also extracted and added to a metadata index that is immediately available to support searching.

## **Indexing the Archive**

The final step in the process of creating a message archive is the generation of the full-text index and compressing the archive repository. As mentioned earlier, the message metadata is immediately populated into the repository's metadata index. The full-text index is generated and updated in a batch process during the Build operation, as the amount of data extracted can be processor intensive. The IMR full text index is extremely efficient, often compressing data at ratios of 20:1 for text-based data. This means that for every megabyte of full text content represented in the archive, the index represents only 50k of data.



Additionally, all messages contained in the archive are compressed at ratios typically 2:1 or greater, further enhancing the compactness of the archive.

The archive repository itself consists of a secure, encrypted data container along with files to manage both metadata indexes as well as the full-text index. These data containers and index files are storage-independent, meaning they can be stored on any storage medium that is addressable from the Archive Service. This is a key feature that enables the ability to store archives on removable media such as compact disc, DVD or the new high-density UDO.

### **Client-Based Access**

Rather than task the IT manager with retrieving messages from the archive, MailStore for Exchange provides the ability to access the archive from an included administrator application or from an Outlook plug-in. This functionality may be the IT manager's "best friend", granting end users permission to search and retrieve specific archived messages that may be needed for compliance reasons without taking additional time from the IT department's typically busy schedule.

This administrator application is different from the MMC program used to create and manage the Archive Service (rules, etc.). This is a Windows client application for records managers, compliance officers, privacy officers, lawyers or other users who require global access to the archives. The app allows the user to do simple or sophisticated metadata index and full-text searches of the entire archive or for specific user directories. It also allows the user to create subsets of archives and even archive to removable media.

There is also an optional Outlook plug-in for simple searches by end users.

### **End-user Archiving**

MailStore for Outlook, an optional extension, provides an easy way for users to add individual messages, message folders, and entire PST archives to the message archive. This client also



allows users to search for and retrieve contents of message archives in a similar manner as the administrator application.

### **Archive Security**

Archive security is maintained on one of two levels. At a minimum, an administrator can apply a database-level or folder/document-level User ID and password directly to the database. One or more ID/PW groups can exist and the combination(s) is actually stored with the database, so it provides a great way to manage access control when it is archived and subsequently accessed by another system using a standard Alchemy client or API-driven access.

Another form of access control integrates with Microsoft Windows security. This integration allows for an IT administrator to match Windows domain users and groups with access levels defined for each database. An administrator can elect to grant access to one message, a collection of messages, an entire archive, or a collection of archives. The Alchemy Server application manages this type of access control.



## **MailStore for Exchange Case Study**

### *Situation*

A leading direct-to-home provider of diabetes and respiratory medications and supplies with hundreds of thousands of customers, many of whom communicate with the company via email, decided it was prudent to keep a record of every internal and external email, and analyze those records for compliance. But with millions of messages a year, this could not be managed on its Microsoft Exchange mail servers.

The company was attempting to meet this objective within Microsoft Exchange, but the procedure was awkward and time-consuming. The email messages occupied about 14 gigabytes of space on the Exchange Server, and short of backing up every individual mailbox, the IT staff had no simple or efficient method to archive and search the messages. It had to back up the entire Exchange database, a procedure that took 8 hours.

### *Solution*

The company invested in MailStore for Exchange. MailStore automatically archives email from 1,600 Microsoft Exchange mailboxes into a secure repository and provides easy retrieval, sorting, and reporting capabilities.

### *Results*

MailStore provides several benefits. It allows specified users to search messages in real-time by metadata or even the full-text contents of messages and attachments. When a senior manager needs to access any email correspondence, he or she can simply enter a keyword and call up all related emails. In addition, MailStore helps the company comply with internal company regulations for email use that protect both employees and the company as a whole. For example, the privacy officer uses the tool to search for unacceptable email correspondence, such as the chain letters that use up valuable bandwidth or messages that might contain offensive language. MailStore has replaced this procedure, freeing up valuable computing resources and employee time.

