

# Advantages of Mimosa NearPoint<sup>™</sup> For E-mail Archival

A White Paper

By Bob Spurzem

January 2005

**Microsoft<sup>®</sup>**  
**GOLD CERTIFIED**  
*Partner*

---

## Contents

|  |    |
|--|----|
| Introduction.....                                | 3  |
| The Demand for E-mail Archival .....             | 3  |
| Regulatory Compliance .....                      | 3  |
| Legal Discovery .....                            | 5  |
| Storage Management .....                         | 5  |
| Challenges of Traditional E-mail Archival .....  | 6  |
| SMTP Data Collection Method .....                | 6  |
| MAPI Data Collection Method.....                 | 6  |
| Microsoft Exchange Journaling.....               | 7  |
| Mimosa NearPoint for E-mail Archival.....        | 9  |
| One Pass Protection™.....                        | 10 |
| A Note about Microsoft Exchange Journaling ..... | 10 |
| Smart Message Extraction .....                   | 10 |
| Self-Service Recovery .....                      | 11 |
| Self-Service Audit.....                          | 13 |
| Security and Authentication .....                | 13 |
| Retention Management and Disposition.....        | 13 |
| Conclusion.....                                  | 14 |

---

## Introduction

Organizations are all faced with the challenge of managing e-mail. E-mail users who send and receive e-mail daily are demanding increased storage space to save months and year's worth of e-mail. Recent federal, state and local regulations recognize the critical nature of e-mail and specify that electronic messaging be managed for privacy and retention. Commercially available e-mail archive solutions perform the basic task of collecting e-mail and storing it in an indexed database for fast search and retrieval. Selected solutions contain features that satisfy specific e-mail archival rules for compliance such as SEC Rule 17a. The data collection methods used by e-mail archival solutions determine the breadth of information and the impact on Exchange. Ideally, all the rich information contained in e-mail is preserved without adding a performance burden on Exchange. This white paper reviews the **Mimosa NearPoint™ for Microsoft® Exchange Server** and compares it to traditional solutions for e-mail archival. Mimosa NearPoint is the industry's first data management solution for Microsoft Exchange that unifies data protection and e-mail archival into a single integrated solution.

## The Demand for E-mail Archival

Despite security concerns, spam, and increasing volume, e-mail is used by 100% of organizations to conduct business. In a recently conducted survey by Kahn Consulting, it reported that organizations are aggressively adopting email for highly-sensitive and valuable business processes and transactions, with 93% using email to answer inquiries from customers, with 84% using it to discuss business strategy, 71% to negotiate contracts, 69% to exchange invoices and payment information, and 44% to file with official bodies.<sup>1</sup> The rapid adoption of e-mail for business communication is forcing organizations to review their e-mail management policies for the purpose of protecting the crucial information they contain and to manage total costs. The demand to archive e-mail is driven by three key areas: regulatory compliance, legal discovery and storage management.

### Regulatory Compliance

Industries that are heavily regulated such as financial services and healthcare must retain electronic messages to meet the regulatory requirements of their industry. The Securities and Exchange Commission (SEC) is the most prominent regulatory body in this regard and requires its broker/dealer members to retain all electronic communication for three years. If your organization is governed by the SEC, then you are already aware of these requirements. However, virtually all organizations are under the governance of local, state or federal regulations that broadly infer that electronic messaging must be retained the same as business records. Broadly speaking, e-mail is used daily to conduct business and contains crucial information

---

<sup>1</sup> "Managing E-mail in the New Business Reality", Kahn Consulting and AIIM International, 2003.

regarding business strategy, contracts, sales invoices and payment information that must be preserved the same as traditional paper letters and invoices.

The Sarbanes-Oxley Act of 2002 (SOX) was signed into law in the wake of the financial collapse of Enron, WorldCom and Tyco. It was designed to improve the quality of financial reporting and included specific fines and prison sentences for failure to comply. Much of Sarbanes-Oxley deals with the processes that involve financial reporting and approval, but a portion of Sarbanes-Oxley deals with audit and review records, including memoranda, correspondence and electronic communications. Sarbanes-Oxley specifies that auditors and accountants must retain this electronic correspondence the same as financial working papers.

| <b>Comparison of Regulations that Impact E-mail Retention</b> |   |  |   |
|---|---|--|---|
|   | <b>Regulator</b>  | <b>Business Type</b>                             | <b>Impact</b>   |
| <b>SEC Rule 17a-3 &amp; 17a-4</b>                             | Securities and Exchange Commission  | Broker/Dealer (brokerage)                        | Specific rules affecting electronic message storage and accessibility.                                    |
| <b>NASD Rules 3010 &amp; 3110</b>                             | National Association of Securities Dealers                                | Securities Dealers                               | Requires specific rules for sampling and reviewing broker/dealer messages.                                |
| <b>HIPAA</b>  | Health Insurance Portability and Accountability Act                       | Health Care Providers; Health Insurance Firms    | Specific rules that cover any type of record that contains Protected Health Information (PHI).            |
| <b>21 CFR Part 11</b>   | Food and Drug Administration  | Pharmaceutical companies                         | Impacts the authenticity, integrity and confidentiality of electronic records.                            |
| <b>DoD 5015.2</b>   | National Archive and Records Administration and the Department of Defense | Companies that sell to the Department of Defense | Specifies that electronic messages must be managed as part of a Records Management Application (RMA).     |
| <b>Sarbanes-Oxley Act</b>                                     | Securities and Exchange Commission  | All public companies                             | Auditors and accountants must retain all electronic correspondence in connection with an audit or review. |

---

## Legal Discovery

In a recent survey conducted by Osterman Research, 71% of users responded favorable when asked, "During the past three years, has your IT department been required to search through backup tapes to retrieve one or more emails in response to a request from legal, HR, etc.?"<sup>2</sup> In other words, three out of four organizations have suffered the arduous task of searching old messaging backup tapes. Recovering e-mail from tape is nearly an impossible task for several reasons. First of all, unless old backup tapes are properly labeled and stored they may be difficult to find and may not be in the best condition. A backup tape is normally restored to a recovery server, not a production server, which is not available to all organizations. The tapes are likely based on different versions of the message server application and this requires that the recovery server be loaded with the matching OS. Once tapes are restored, their content must be searched and is limited by the native messaging search tools. A common practice is to print out e-mail for manual review. The bottom line is that backup tapes are an expensive and impractical method of preserving e-mail for archival. The technical difficulties and time necessary to search and retrieve historical e-mail stored on backup tapes are just too great to make this method feasible.

## Storage Management

The rapid adoption of e-mail and the corresponding demand for CPU and storage resources has put a tremendous strain on IT environments and budgets; an issue that every IT Administrator is painfully aware of. If the e-mail content is only valued short term, then the problem is not too painful. E-mail quotas can be set to low levels (e.g. 10MB) to allow day-to-day communication with no long term retention. But typically, the information that e-mail contains is crucial to business operations and is necessary for later recall. Organizations that use e-mail quotas have observed that users will circumvent quotas by storing e-mail locally in Personal Store (PST) files. Organizations that allow users to store e-mail for long term are experiencing the constant (and costly) demand for additional storage resources. Osterman Research surveyed a large number of organizations and found message storage is growing an average of 24% per year.<sup>3</sup> In this same survey, 41% responded that individual mailboxes exceeded 200MB in size. Whether e-mail is stored locally in a PST file or stored on the message server, the trend is clear – users (and businesses) demand long term e-mail retention.

---

<sup>2</sup> "How to Evaluate and Choose a Messaging Archiving Solution", Osterman Research Inc. and Contoural Inc., page 9, 2004.

<sup>3</sup> IBID, Osterman Research Inc. and Contoural Inc., page 7, 2004.

---

## Challenges of Traditional E-mail Archival

E-mail Archival applications are available commercially and perform the task of preserving electronic messaging long term. The basic features that these solutions provide are an indexed database to catalog records, search and retrieval, and security. They benefit messaging servers by reducing the burden of managing old e-mail and allowing archival access for search and retrieval. Highly regulated industries such as the SEC have specific requirements that drive the selection of messaging archival solutions. However, other industries lack specific technical requirements for archival and allow a broader selection of products. Typical comparisons of messaging archiving products delve into the details of storage, search and security and pay little attention to the methods which data is collected from the messaging servers. The data collection methods vary by the richness of message information they collect and depending on the method, can severely impact Exchange performance.

### SMTP Data Collection Method

“SMTP” stands for “Simple Mail Transfer Protocol” and is the messaging protocol used between messaging servers. SMTP Data Collection is the method of intercepting e-mail at the gateway server, while the message is in its SMTP format. This method has the advantage of collecting e-mail at a single point; a method that is less complex than collecting e-mail from multiple e-mail servers. And because SMTP is compatible with all message servers, all the popular e-mail types (e.g. Microsoft Exchange, IBM Lotus Notes, etc.) are supported. All data collection processing is performed on the gateway and does not burden the messaging server(s). SMTP Data Collection is a favorite method of e-mail archival service providers.

The drawback of SMTP Data Collection is that it is either limited to inbound/outbound messages only, or it forces all internal messages to be also relayed through the gateway, thereby increasing the load on the messaging server. SMTP also restricts the rich content capabilities provided by messaging systems and contains only basic information for date, time, from, to, cc, subject, body and attachment. Beyond this basic content information, there exists rich information that relates to the context and lifecycle of the message. Message context includes its folder location, flags, rich text and settings. Message lifecycle refers to movements including replies, forwards, edits, opens, deletes and folder changes. Messages viewed at the SMTP gateway, lack this context and lifecycle information. They only contain basic message content information and lack the rich information provided by messaging systems.

### MAPI Data Collection Method

“MAPI” is the acronym for the Microsoft Message API and is the Microsoft supported method for reading messages in Microsoft Exchange Server. Traditional e-mail archive solutions use MAPI to scan messages in the Message Store for archival. Following e-mail archive policies, messages are copied to the archive

---

and indexed for fast search and retrieval. Optionally, messages are removed from the production Exchange server and replaced with a small "short cut". This benefits the production Exchange server by reducing the burden of storing old e-mail records.

The major drawback of MAPI is the performance burden it places on the Exchange server. To lessen this burden, e-mail archival solutions limit the use of MAPI in two ways. First they restrict the amount of information collected by MAPI for each message. Each message contains its header, body and attachment (if one exists) and over 400 individual properties. These properties contain the message context and lifecycle information of each message. Because of the time it takes to read the message content, the message properties are not collected. This reduces the overall burden of MAPI on the Exchange server, but limits the amount of information available for regulatory and legal discovery analysis.

The second method that is used to limit the burden of MAPI on the Exchange server is to schedule MAPI on the weekend when e-mail use is low. This MAPI schedule limits its view of the Message Store to once a week. Messages sent, received and deleted (and purged) during the week are not available and therefore not available for archival. For organizations whose objective is to archive old e-mail for storage management purposes this method works well. Only messages that exceed a certain age (e.g. 60 days) and still remain in mailboxes are copied to the archive server. But organizations that archive e-mail for compliance and must retain "all e-mail", require the additional services of Exchange "Journaling".

## **Microsoft Exchange Journaling**

Microsoft Exchange includes a standard server feature that archives all messages sent and received by mailboxes on any given Message Store. This feature is commonly referred to as "Journaling". Journaling consists of a duplicate Message Store located on the Exchange Server. MAPI Data Collection methods that typically run on a weekly basis use Journaling to capture e-mail traffic during the week. The drawback of Journaling is the CPU and storage burden it puts on Exchange. Journaling captures one additional copy for each message sent or received by a server from outside. It forces even internal messages to go through the expensive route that outbound messages take even if the recipients are on the same server. Exchange Administrators are wary of Journaling for this reason. To manage Journaling for compliance, E-mail archive solutions use MAPI to read the journal mailboxes and truncate its contents. Without MAPI, the Journaling mailboxes would quickly fill and consume valuable Exchange storage space. The combination of MAPI and Journaling doubles the performance burden on Exchange.

The second drawback of Journaling is its lack of rich message information. The journal mailbox is a stripped down version that bares no similarity to the assigned mailbox except for the content of the messages. This limits the amount of message context information available when MAPI reads the journal mailbox. Journaling only views the message when it is sent or received; otherwise it loses all contact with the message and its history. All the rich lifecycle information contained in the user's mailbox (open, replies, forwards, etc.) is not available for journaling.

Journaling is required by e-mail archive solutions when the objective is to capture 100% of e-mail for compliance. The drawback of Journaling is that it places a significant performance burden on Exchange and it does not contain the rich message information that is associated with e-mail when it exists in normal mailboxes.

| <b>Comparison of E-mail Archive Data Collection Methods</b> |                     |                             |                    |
|---|---------------------|-----------------------------|--------------------|
|   | <b>SMTP</b>         | <b>MAPI</b>                 | <b>Journaling</b>  |
| <b>Collection Point</b>                                     | Gateway             | Exchange Server(s)          | Exchange Server(s) |
| <b>Collection Process</b>                                   | Continuous          | Point-in-time (e.g. weekly) | Continuous         |
| <b>Performance Impact on Gateway</b>                        | High                | None                        | None               |
| <b>Performance Impact on Exchange</b>                       | Medium <sup>4</sup> | High                        | High               |
| <b>Storage Impact on Exchange</b>                           | None                | None                        | High               |
| <b>Richness of Message Information</b>                      | Minimum             | Low                         | Low                |
| <b>Message Servers Supported</b>                            | All                 | Exchange Only               | Exchange Only      |

<sup>4</sup> If internal messages are to be archived then every message being sent and received must go to an extra destination and increases the performance burden on Exchange.

# Mimosa NearPoint for E-mail Archival

**Mimosa NearPoint for Microsoft Exchange Server** is the industry's only comprehensive data management solution for Microsoft Exchange that unifies data protection and e-mail archival in a single integrated solution. (figure 1.) NearPoint solves many of the challenges facing Exchange data protection, but this paper focuses on the e-mail archive capabilities of NearPoint. The Mimosa engineers designed NearPoint to overcome the drawbacks and limitations of traditional e-mail archive solutions. Two major design goals were to perform all processing "off-host" and to gather as much rich message information as possible. In its initial delivery, Mimosa NearPoint for Microsoft Exchange Server performs all the typical functions of e-mail archival "off-host" without adding any performance burden to Exchange. The e-mail archive functions that NearPoint performs are indexing, full-text search, authentication, retention, disposition, and security. In addition, NearPoint provides a complete view of mailboxes, including folders, contacts and calendars and it can re-construct complex message threads and conversations across time and across multiple servers. The following sections describe in detail the important functions that Mimosa NearPoint performs for e-mail archival.

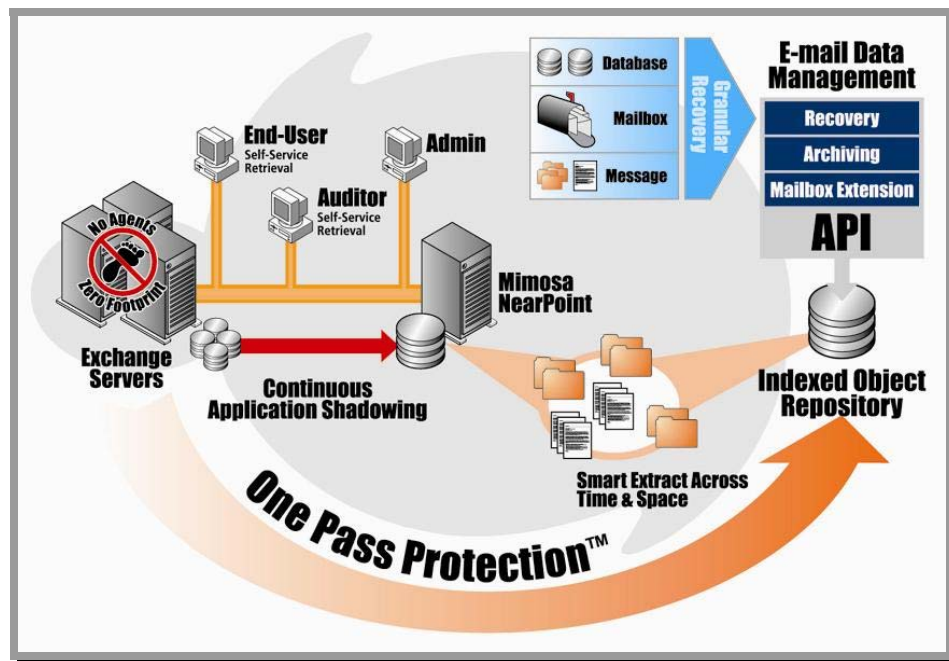


Figure 1. Mimosa NearPoint Architecture

---

## One Pass Protection™

Mimosa NearPoint does not rely on SMTP, MAPI or Microsoft Exchange Journaling to gather messages for archival; rather it uses a proprietary method called One Pass Protection™. One Pass Protection performs near continuous data collection of messages for archival and uses the Exchange transaction log files as the source of rich message information. After a full Exchange database backup, using the Microsoft Exchange Extensible Storage Engine (ESE) Backup API, One Pass Protection continuously copies the transaction log files to the NearPoint server where they are applied to the full Exchange database copy. This log shipping process is called “Continuous Application Shadowing” and it is a superior method of capturing rich message information as compared to the SMTP and MAPI methods. The amount of processing required to read the transaction log files from Exchange is minimal and the information contained in the log files is rich with message context and lifecycle information. In contrast, the SMTP and MAPI methods both lack rich message context and lifecycle information and MAPI places a significant performance burden on Exchange as previously discussed.

### A Note about Microsoft Exchange Journaling

Mimosa NearPoint supports Microsoft Exchange Journaling and One Pass Protection can be configured to read journal mailboxes. NearPoint supports Journaling for customers who are accustomed to its use; but this is not a requirement. Mimosa recommends the use of One Pass Protection without Journaling because it efficiently captures all message information for archival from log files, making Journaling unnecessary. The Exchange transaction log files are designed for Exchange data protection and are an ideal source of rich message information for archival. They contain 100% of the message content (e.g. header, body and attachment) and they contain all the message context and lifecycle information. One Pass Protection efficiently copies the log files to the NearPoint server and in the process captures all of the rich message information for archival. Traditional e-mail archive solutions that rely on MAPI and Journaling do not collect all the rich message information important for archival analysis and the collection process they use is slow and burdens Exchange.

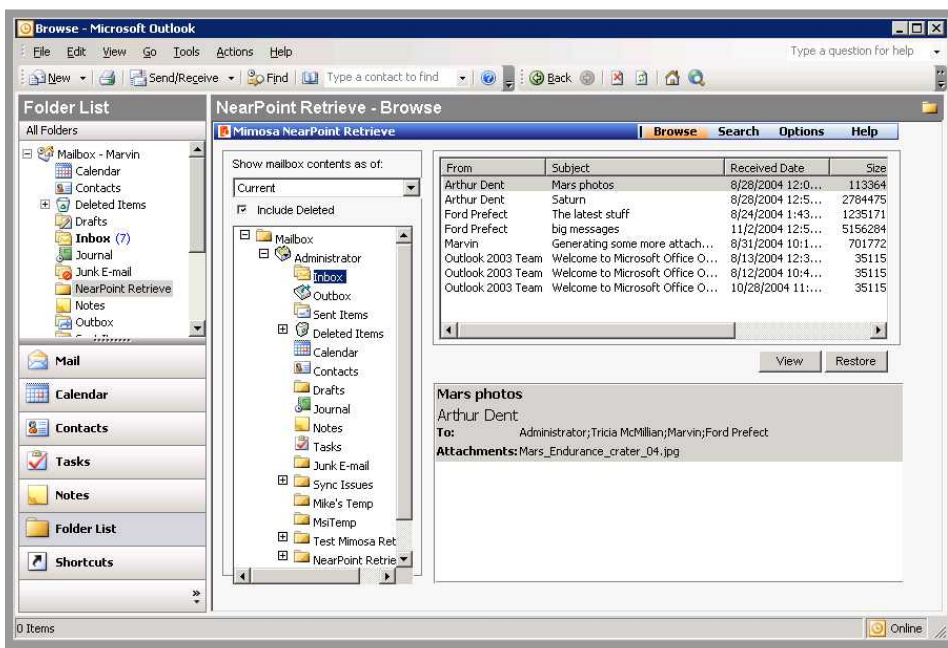
### Smart Message Extraction

Smart Message Extraction is the name of the method that NearPoint uses to process the Exchange message data for archival. Smart Message Extraction runs entirely on the NearPoint server and places no burden on Exchange. Each time transaction log files are received on NearPoint and applied to the full Exchange database replica, Smart Message Extraction automatically runs and processes the data for archival. The first step of the Smart Message Extraction process is to break each individual message into its major components – header, body and attachment (if one exists). Each component is indexed, compressed and stored in NTFS. Next, Smart Message Extraction catalogs the property information for each

message and stores it in a Microsoft SQL Server Desktop Engine (MSDE) or can leverage an existing Microsoft SQL Server database. This rich message information includes context information for folders, permissions, flags and rich text and it includes a record if the message was opened, edited, replied, forwarded or deleted. Mimosa calls the database that stores all this metadata information the Indexed Object Repository. The Indexed Object Repository is the “heart” of the e-mail archive. Using the information it contains, it is possible to quickly search the archive and view complete mailboxes at any point-in-time.

## Self-Service Recovery

End users, Auditors and Message Administrators all have self-service access to the Indexed Object Repository to perform full-text search of messages and browse mailboxes at any point-in-time. Access for search and browse is via the standard Microsoft Outlook and Outlook Web Access (OWA) interfaces and a folder that NearPoint publishes on the Exchange server called the “NearPoint Retrieve” folder. When the NearPoint Retrieve folder first opens, it displays a browse view of the entire mailbox at the current time. (figure 2.)

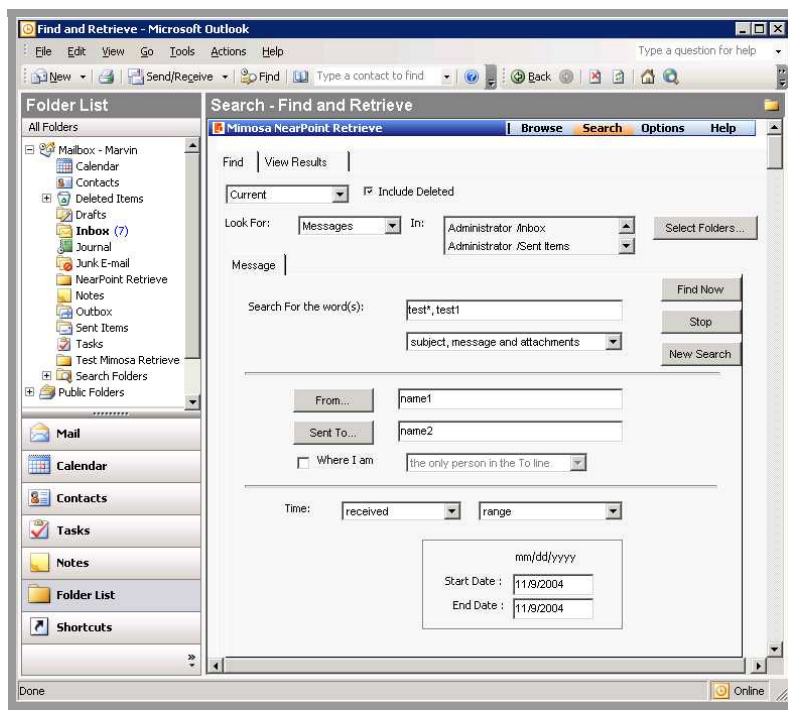


**Figure 2. NearPoint Self-Service Retrieve “Browse Screen”**

The NearPoint browse screen consists of three displays. The complete mailbox including all the mailbox folders, Sent Items, Drafts, Notes, Deleted Items, Contacts and Calendar is displayed on the left side pane. For the folder selected, contents are displayed in the top pane and a preview pane is located immediately below. An individual message can be viewed with a simple “double-click” just like standard Outlook. The browse window time defaults to current time, but it can be changed to the time you select. Mimosa NearPoint also provides a search screen for full-text searches of the archive. (figure 3.) Using the search fields, it is easy to

search messages by time, by sender, by recipient and by message text. Information contained in the message header (e.g. From, To, CC, Title), message body and attachment (if one exists) is indexed for fast search and retrieval. Search results are displayed on the screen and can be saved to a folder or exported in a PST file. Searches themselves can be named and saved for later use. This feature is useful if a custom search must be repeated.

Self-service recovery using Mimosa NearPoint is more powerful than traditional e-mail archival solutions because it can view the full context (e.g. folder location, permissions, properties) and lifecycle of mailbox and the messages it contains. Using NearPoint, Auditors can browse a mailbox at a specific point-in-time and uncover the full relationship between messages, folders and actions taken by the mailbox owner. Say for example, questionable activity in a certain mailbox occurred one month ago in the early evening. With a few simple clicks, the full mailbox at the time in question can be displayed using the NearPoint browse screen. Auditors can open folders and individual messages to view their full contents and context. If a message was opened, edited, forwarded or deleted, this information is also available. In this manner, the complete relationship of the mailbox and its contents can be ascertained for the purposes of the investigation. In matters of legal discovery and investigation, it is valuable to view not only the contents of the message but also its surrounding mailbox to determine the true intent of the user whose mailbox is being investigated.



**Figure 3: NearPoint Retrieve Folder Search Interface**

---

## Self-Service Audit

Using NearPoint, auditors can be granted access to multiple mailboxes for self-service search and retrieval. This saves time and avoids any privacy issues that are implicit when searching other people's e-mail. The auditor views these mailboxes using the standard Outlook and Outlook Web Access (OWA) interfaces and uses the same browse and search screens as end users; but differs by the ability to access more than one mailbox. Powerful search functions allow auditors to look for emails as well as other items including calendar events, contacts, and keywords within subject line, message body, and/or attachments. Searching the "From" and "To" fields as well as defining a time period for a search allows for further granularity of detail, increasing the efficiency of the discovery process and allowing the ability to recreate a point-in-time view of communications between individuals.

## Security and Authentication

Mimosa NearPoint protects the integrity of the Exchange archive information in two ways. First access to the archive is only provided to authorized individuals. The Exchange Administrator holds all the "keys" and is responsible for granting access to the proper individuals and managing their access. Secondly, Smart Message Extraction computes a digital signature for each object (e.g. message header, body, attachment) using the MD5 Hashing Algorithm, when it is received in the archive. The digital signature is stored in the MSDE or MS SQL database along with the object and is used in future dates to verify the integrity of the data. The digital signature is also compared to existing records and if a match is found, the object is marked as a duplicate and only a single copy of the object is stored. This feature is referred to Single Instance Storage. Along with standard data compression on each object, the data in the NearPoint Indexed Object Repository remains tamperproof and its integrity is verifiable.

## Retention Management and Disposition

Mimosa NearPoint has two retention management policies that are configured by the Exchange Administrator. The first retention management policy manages all records in the archive based on the data protection period. This policy disposes of records a period of time (e.g. 30 days) after the data is deleted on Exchange. This policy is useful for removing data which has been deleted on Exchange and is not required in the archive long term. The second retention policy manages all records in the archive based on the archiving retention period. This policy disposes records after a period of time (e.g. years) has expired based on the creation date. This policy is used to manage a company e-mail policy that stipulates all e-mail be retained for a period of time and then disposed. Both methods can be set independently at the mailbox, storage group, or Exchange level. If an investigation is taking place, the disposition processes can be suspended.

---

## Conclusion

**Mimosa NearPoint for Microsoft Exchange Server** is the next generation data management solution for Microsoft Exchange that unifies data protection and e-mail archival into a single integrated solution. As an e-mail archive solution, NearPoint uniquely performs Exchange data collection using a method that avoids the most common problems associated with the MAPI and SMTP data collection methods. By leveraging the Exchange transaction log files, NearPoint is able to provide full message content, context and lifecycle information without burdening Exchange. Smart Message Extraction runs entirely "off-host" and processes the Exchange data for archival. The indexing, hashing and compressing processes that Smart Message Extraction performs enables self-service archival access by users and auditors. Using NearPoint self-service retrieval, auditors can re-construct complex message threads and view entire mailboxes at a point-in-time. The archive records are protected against false tampering and their integrity is verifiable. Archive storage is optimized by the global single instancing of message bodies and attachments. Retention policies manage records according to their retention policy and dispose of them when the policy expires.

For more information about Mimosa Systems and its new NearPoint solution, contact your Mimosa Sales Representative at 1-(408) 970-9070 or visit our web site at [www.mimosasystems.com](http://www.mimosasystems.com).

### ABOUT MIMOSA SYSTEMS

Mimosa Systems™ provides immediacy, discovery & continuity for enterprise information. We combine fine-grained, immediate recovery with self-service archival access for messaging and similar data, leveraging cost-effective disk technologies. Mimosa NearPoint™ for Microsoft Exchange unifies data protection, mailbox extension and archiving in a single solution, assuring email continuity and regulatory compliance.

Mimosa Systems, Inc. | 3200 Coronado Drive, Santa Clara, CA 95054 | [www.mimosasystems.com](http://www.mimosasystems.com) | 408-970-9070

Copyright © 2005 Mimosa Systems, Inc. All rights reserved. Mimosa, Mimosa Systems, Mimosa NearPoint, NearPoint, One Pass Protection and NearPoint Deployment Advantage are trademarks of Mimosa Systems, Inc., in the U.S. All other company and product names are trademarks or registered trademarks of their respective companies.