



Gone in 30 Days: Exchange, Retention, and Regulatory Compliance

Paul Robichaux

Principal Engineer

3sharp

MCSE, and an Exchange MVP

Balaji Srinivasan

Senior Product Manager

Sherpa Software

The advent of Sarbanes-Oxley, Gramm-Leach-Bliley, and assorted market-specific regulations means that you may be legally required to have an email compliance and retention policy. Although we can't tell you what that policy might be, in this white paper, we'll share with you an overview of general retention and compliance issues, an understanding of Microsoft Exchange Server's built-in archiving and compliance features, knowledge of some sneaky pitfalls that may await you in implementing your policy, insight into how to manage your mail data for best-efforts compliance, and guidance on first steps to take when starting an archiving regime.

Also, to satisfy corporate, legal, or regulatory requirements, organizations need a way to locate their critical data within the email and move that data to a secure location where it can be protected long-term. In this white paper, we will discuss how to discover, manage, and archive information within your Exchange enterprise to successfully limit your legal exposure and protect your corporate information. We'll explain how you can analyze trends and usage across your entire messaging store; implement retention policies in Exchange mailboxes, PST files (network/local), and public folders; search and retrieve specific content for the archive; overcome the 'PST Pitfall'; understand archiving options and Exchange limitations, and manage the archive.

→ Contents

Retention and Compliance Policies	1
The Differences Between Archiving, Retention, and Compliance	2
Key Questions to Ask	3
Basic Retention Requirements	4
What Exchange Can Do Out of the Box	5
Retention and Compliance Policy Pitfalls	6
Best Efforts Compliance	7
Addressing Email Storage Concerns	8
Mail Storage Best Practices	9
<i>Sherpa Mail Attender</i>	9
<i>Tackling PST Files</i>	10
Managing the Data You Keep	12
<i>Storage Archiving</i>	12
<i>Compliance Archiving</i>	13
<i>Targeting Content to Be Archived</i> ..	13
<i>The Information Life Cycle</i>	13
Sherpa Solutions for Email Management	14
About the Authors	14

sponsored by Sherpa Software



Copyright 2005 Windows IT Pro. All rights reserved

Gone in 30 Days: Exchange, Retention, and Regulatory Compliance

Retention and Compliance Policies

What does it mean to implement retention and compliance policies with Exchange? Very briefly, there are five main points that we want to cover:

1. The difference between retention and compliance
2. What kind of features Exchange provides out of the box for retention and compliance in implementations
3. Some of the pitfalls that you can get into with Exchange retention and compliance policies
4. What it means to undertake a best-efforts compliance policy
5. A discussion of some of the first or early steps you can take to begin making your organization compliant

Before we get started, let's include a short disclaimer: We're not attorneys, we do not pretend to be, and we're not legal experts. In the Exchange Security book *Secure Messaging with Microsoft Exchange Server 2003* (Microsoft Press) that Paul Robichaux wrote earlier this year, he actually hired a lawyer to write the messaging law chapter. So we're not going to be giving you legal advice. Ordinarily we would not bother to say that, but many organizations are very sensitive about what their retention and compliance implementations look like because they want to make sure they are meeting the legal requirements that are imposed on them. So if you are trying to implement this policy, and you do not have anyone in

your organization who can guide you through what your actual legal requirements are, and to help you make sure that your messaging system design will meet those requirements, then you should probably import or hire out that particular role, to make sure that you are getting good legal guidance.

The Differences Between Archiving, Retention, and Compliance

First, it helps to understand exactly what we're talking about, and what the differences are, between retention and compliance. Every aspect of the law has its own specialized vocabulary, and this is really no exception.

For the purposes of this white paper, *archiving* means keeping mail and messaging data around for a defined period. Depending on your organization, depending on what it is you do—if you are a state government, a local government, part of the federal government, a university, a business, a publicly traded company, and so on—you may have requirements to keep your mail data around for a certain period of time—say 7 years, 7 years, 2 years, 90 days. The exact period will vary between organizations, but the goal of implementing archiving is to be able to keep your mail around for as long as you need to, so that you can search it, process it, migrate it between storage sites, and then get rid of it when you do not need it any more.

Retention is archiving plus, if you will. It refers to an archival process that lets you take your mail and keep it for a defined period, while still maintaining good control over who has access to the archived mail, the integrity of the archived mail—a chain of evidence. In other words, you should be able to show exactly what has happened to a message at every given step in its lifetime, in such a way that it would be legally admissible, if necessary—and then reporting, so that you can show that you are actually retaining the messages you are supposed to retain, and keeping them as long as you are supposed to keep them, and finding the correct number of messages when you do test queries, and so forth.

Compliance means using archiving and retention to meet a defined set of legal or regulatory requirements, and then being able to prove that you actually meet those requirements. The simplest case of compliance is complying with a court subpoena. So the FBI shows up at your door with a piece of paper that says, “We are subpoenaing all your records pertinent to your dealings with Enron,” or “...pertinent to your dealings with this individual.” At that

point, you are legally compelled to produce those records. But it is also important that you be able to *prove* that you actually have produced all the records that you have. And so having a complete compliance solution means that you are including the ability to do that kind of reporting.

We want to mention, just as a side note, that earlier this year an analyst firm put out a study that said that about 7 percent of the United States' publicly traded companies had not yet complied with Sarbanes-Oxley. Now, we found that very surprising, given how long people have had to get Sarbanes-Oxley compliance in place, and what the financial and legal penalties are for not doing so. But even so, we think if you were to look at the number of organizations that have to *comply* with SarbOx, or with HIPAA, or with the EU Data Protection Act, or with other similar regulatory regimes, and if you took the number of companies in total that are subject to those regulations, you would find a significantly higher percentage than 7 percent of companies who *think* that they are compliant but really are not. Now, those are just the legally compelled compliance frameworks. Over and above those, in some industries there are regulations or practices that govern how you do retention, what kinds of things you keep, and for how long, and so forth. Anyone from a law firm will probably have a pretty good idea of exactly what we are talking about. The American Bar Association has standards that they are promulgating that explain, or that set forth, how long you may keep things around, how long you must keep things, and so forth.

Any organization that has a legal department is liable to be getting guidance from their own lawyers who talk about the specific implementations of compliance and retention that make sense in their individual environment. The requirements for a very large, publicly traded company (e.g., Hewlett-Packard, Chevron, Boeing) are going to be different from those requirements for smaller companies, just because of the size of the company. By the same token, in some industries, or in some sectors, there will be more onerous requirements. If you were a consumer-products company, for example, you would expect to have a more stringent retention regime than if you were a manufacturing company that made, say, steel frames for furniture. Anything that touches on any potential litigation or sources of litigation is going to be an area where you will see internal legal guidance setting a more restrictive or more demanding standard for compliance and retention.

Going back to the issue we mentioned earlier, of subpoenas, consider the instructions from civil or criminal cases. In some cases, companies found out, to their dismay,

that they were not able to produce everything called for in a subpoena. So those of you who are frequent watchers of what Microsoft is doing or has done in court, you may remember that earlier this year it settled a dispute it had with a small company called Burst. The dispute was over whether Microsoft had stolen some technology from Burst. And during the trial, the Microsoft attorneys said that they had produced all the email that was respondent to a subpoena that Burst submitted. As it turned out, Burst was able to produce copies of email that was not included in Microsoft's discovery delivery. In other words, Burst had copies of emails that Microsoft sent that did not appear in what Microsoft said was a total and complete record of all the mail they had. It very probably contributed, in a non-legal opinion, to Microsoft deciding to settle the case because, if the case had proceeded to trial, the fact that Microsoft could not produce those messages certainly did look suspicious, even though it claimed to have done nothing wrong.

Key Questions to Ask

As you assess what kind of compliance setup makes sense for you, there are a few questions that you should ask.

What to retain. The first question you should ask is, "What do you have to retain?" Now, this might seem like an odd question because ordinary retention wisdom is that you should retain everything. But you might not need to retain everything. Clearly, there is no good legal reason to keep around things like nondelivery receipts, or message-delivery reports, or messages generated by the mailbox-manager process. There may be exemptions, or there may be opportunities for you to not retain things that are companywide announcements: For example, "There is a blue Honda in the parking lot with its lights on"—those types of incidental messages that do not actually concern the company's business. You may also have mail that you *must* retain, based on who generates it. Now, this point is a little sticky, but at many publicly traded companies, the retention requirements are most strictly applied to people who are directly involved in the high-level functioning of the company; so the CEO, the Chief Financial Officer, members of the Board of Directors if they have company email boxes, and so forth. Setting the scope of a policy by deciding who is and who is not included can be very valuable because it gives you a bright line, if you will, that you can draw that says, "Okay; these people are generating ordinary work day mail; I am not going to include them. These people at a higher level are more likely to be knowl-

edgeable about deep internal affairs, if you will, so we are going to ensure compliance by making sure that we are keeping track of their mail."

How long to retain. The next question to ponder is, "How long do you have to keep the mail?" Now, there are basically three choices—or four, really. The first is "Forever." Now in most organizations, that's not tenable. It is possible for you to archive mail, first to disk, then to tape, then to optical media, then to offsite storage. It is even possible to take all of your email and print it and put it in a giant storage vault somewhere—not necessarily a *good* idea, but it can be done. However, what you may find is that your organization adopts a policy that says, for example, "All of the CEOs' email we're going to keep forever and ever, and then we will taper down from there for directors and other officers who have material financial input into the business. We will keep their email for 5 years. And for people between these two pay grades, we will keep their email for 3 years, and for everybody else, we will keep their email for 1 year." It is possible to set up these kinds of graduated policies *if* you have the right kind of retention and archival tools. It is much more common to see organizations setting retention policies that say they will uniformly keep mail for a certain number of days—anywhere between 30 to, say, 730—2 years' worth. The exact number of days that you store email will depend on the legal requirements, but also on how much money you are willing to spend on your archival solution. Because, as your mail volume increases, as it has pretty steadily, you will find that the storage costs for that mail are going to increase also. Even though storage acquisition costs keep dropping, the cost of managing that storage once you have bought it is not dropping nearly as fast. In fact, it is probably flat, and it is showing slight indications of trending upward.

The other two alternatives are probably not going to be suitable for most folks. "Until told otherwise" just means you are going to keep all your mail around until you do not have to keep it any more. This practice is typically what happens when you get involved in a civil or criminal investigation or litigation. When you are a party to a lawsuit, for example, one of the first things that you get is a discovery request from the opposing side that says, "Here are all the things that we want you to produce—all the business records, all the pieces of information, and so forth." From that point forward, until the conclusion of that legal action, it is a really bad idea for you to throw away any documents or records that pertain to that matter. And so you have got to be real careful to make sure that you are not

removing mail from archival that you should keep. Besides “Until told otherwise,” one additional potential retention time is “Never,” in which you do not retain anything. Clearly, that’s not realistic for most organizations. If you say outright that your explicit policy is that “We do not retain any email at any time,” then you are going to have a hard time convincing people not to file things away on their own. And as soon as that starts happening, you open yourself up for a great deal of trouble, as soon as employees are free to individually choose what they keep and do not keep. Because then you have no way to ensure broad compliance across the board.

How to retain mail. The next question is, “How do I retain that mail?” The three primary choices here are terms that will be familiar to everybody who is ever worked as an Exchange administrator. *Online* means you will keep the mail immediately accessible, possibly on your Exchange server, possibly on an archiving server or servers, but it will be online, so people can get to it more or less immediately. *Nearline* means that you will keep it in a storage system where an index is made available for online users, and you can quickly retrieve the message from nearline storage, perhaps using an HSN system. *Offline* means that your archives are kept somewhere where they are not immediately available. You may or may not have a message catalog that can be searched. More likely, though, if you have moved to offline for a portion of your retention, then a request for documents in that retention period will require you to go get your offline archives, make them available by mounting them on a server, and then search them, or doing whatever you have to do to produce the demanded records.

How to prove your retention policy. The *big* question is, “How do you *prove* that you followed whatever your policy is?” This is really the key, sticky point. It is good to have a policy; it is better to be able to show that you are following a policy by producing things like test runs. You say, “All right; on January 1, I did a test run to see how many messages that contain a special code phrase I found.” And then when you do it again the next January 1, you should find the same number of messages. Of course, what we’re actually talking about is using messages that have an unusual or unlikely phrase in them that you put into the messaging system just to test your retention. You might send out 100 copies of a message that say, “Do not delete this,” and the special phrase is “Philadelphia Eagles Win World Series Cup,” to mix a sports metaphor. When you search for that exact phrase, if you sent out 100 copies of that message, you

should find 100 copies. If you find more than that or fewer than that, that tells you that your archiving system is possibly missing mail, or that people are messing with mail that you asked them to leave alone, either one of which is something that you are going to want to pay some attention to.

Basic Retention Requirements

Let us look at the basic requirements for retention. When you start implementing retention, obviously you have to start somewhere. And so the question becomes “Where am I going to start? How am I going to go about putting this strategy into place, to keep me out of jail?”

The first step is to get all your existing mail into your retention system—all the mail from all your Exchange servers, from PST files on people’s desktops, on their laptops, from everywhere. If you do not do that, then you run the risk of the same situation that Microsoft found itself in, where you have got mail that was not included in archival, but which still could be produced at an inopportune time for you. Your retention policy has got to be comprehensive and include every potential storage location.

On an ongoing basis, once you have got that mail in, you have to make sure that you retain inbound and outbound messages that are newly created after the retention start date. You have to make sure that you can generate reports that show that you are being compliant with whatever your retention requirements are. If your inbound message statistics show that you got 5000 messages in a given month from outside recipients, then you should be able to account for all of those. Now, clearly, that puts a premium on your spam filtering and on other aspects of your messaging system because you certainly do not want to have to keep at least the kind of spams we get—you do not want to have to keep those for 5 or 7 years! But once a message enters the retention system, the only way for it to leave is to be purged or removed according to what your retention policy specifies. If your policy specifies that a set end date is possible, then you will have to make a provision to clean out those messages at the end of their lifetime, after the 30 days, or 90 days, or 3 years, or whatever the period may be, has passed. At that point, it will be safe for you to remove those messages, but *only* those messages, from your retention system. This practice typically has to be a rolling process because if you have a message that was created on January 1, 2001, and you keep it for 4 years, you want to remove it at the beginning of January 2005. You do not want to keep it around and do a single, end-of-the-year cleanout. Although you might be able to get by with doing purges quarterly or biannually, depending on how many messages

you have, and on how important it is that you not keep unwanted, end-of-life messages around.

On-demand, something you will only have to do when circumstances dictate, you will need to be able to pull messages that meet certain criteria. For example, Paul once worked with a law firm that had been retained by Enron at one point. This is actually well before Enron became famous as an example of corporate misgovernance. But this law firm, through no fault of its own, was served a subpoena that said “We want to see all the messages you have that have anything to do with Enron, or that contain any of this list of key phrases.” So that put them in some situation because, at the time, they did not have any kind of retention system or policy in place. Once they got that system in place, it was trivial for them to do a keyword search for Enron, and then look over the messages they retrieved, and find out if there were any additional criteria that they needed to use to broaden their scope. In other words, if they had people who were working only on the Enron case, it might be possible that they were exchanging messages among themselves, or with people from Enron, without actually using the word *Enron*. So they were able to include individuals as part of that retention request, to make sure that they were retrieving *all* the messages that were appropriate to respond to that requirement.

What Exchange Can Do Out of the Box

Everything about Exchange described in this section applies to Exchange 2000, Service Pack 3, and later, and Exchange 2003 RTM and later. If you are running Exchange 5.5, you will find it has very limited support for even the limited feature set that Exchange 2000 and later implement. Exchange does not support importing existing mail into an archiving system. It does not have any support for reporting. It does not have any support for purging messages after they have reached their end of life. There is no way to script it, and to automate it, and say, “I want to expire these messages at a particular time.” There is also no support for either keyword or time-based reporting or retrieval—this is something that Microsoft has not implemented, for reasons we do not understand. The ability to do a quick keyword search of selected mailboxes is extremely valuable.

Exchange does offer message journaling. Journaling itself is not a retention or archiving feature; it just refers to Exchange’s ability to capture copies of messages that are either sent or received by users in mailbox databases for which journaling has been enabled. For example, you can

turn on journaling for mail databases and all the users whose mailboxes are in that database will be journaled. Exchange does not let you journal or copy mail for only one user, unless that user is the only user in a mailbox database. When messages are received for delivery, when a user submits a message to the Information Store (IS), it will be received by the IS as though it had been submitted from a remote system. An Exchange component called the Categorizer performs an operation known as *bifurcation*. That term basically means that Exchange will clone the message: One copy will go on its normal way, to its intended destination, and the other copy will be kept by the journaling system.

Now, with Exchange 2000 SP3 and later, or any version of Exchange 2003, three modes exist in which you can operate journaling. In *standard mode*, you only get messages. That means that you do not get nondelivery reports, you do not get read receipts or S/MIME receipts, and you do not get copies of messages that are sent to blind-carbon-copy recipients. And normally that is what you would expect in a messaging system. When you send a message to someone who is a bcc recipient, the whole point of using bccs is that the person who is receiving the message does not see the existence of the bcc recipients. The problem you run into from a retention standpoint is that if you do not take some action to capture bccs, people will be able to use them to evade the journaling. So in *bcc mode*, which is the next step up from standard mode, the journaling system will capture everything that standard captures, plus bcc recipients. The limitation of that approach is that it does not capture recipients who are on distribution lists. This restriction is significant because your distribution-list membership may be significant in doing recovery requests, so that when you are asked to produce all messages sent through a particular user, if all you can do is query based on the distribution-list name, which is what you will get with either standard mode or bcc mode, then you do not have any way to know that Paul Robichaux is a member of a particular distribution list. So you will have to manually go back and look at which distribution lists Paul’s in. And even that is not going to help you figure out what distribution lists he was on 6 months ago, 1 year ago, or 3 years ago.

So when you turn on *envelope journaling mode*, *all* of the recipient information is captured, including bccs and the distribution list recipients to which the message was addressed. Exchange does this by capturing the message after the distribution list expansion has been done on the

distribution-list expansion server. As a result, you will see a message that, instead of being sent to all employees, let us say, will have the list of individual mailboxes in the To: header. Envelope journaling also captures final recipients. Regarding the latter, if you are using address rewriting or if you have SMTP contacts enabled, anything that will modify the initial recipient address into some other form will be captured by envelope journaling. It is an open question of whether you need to keep those, but it is better to capture them and purge them later than it is not to get them in the first place.

Now, as we mentioned, you have fairly coarse-grained control over message journaling. You can enable journaling on an individual database or you can create an Exchange system policy for mailbox databases that will turn on journaling. But you cannot journal individual users, you cannot journal Active Directory (AD) groups, organizational units (OUs), or domains. And it is not trivial to set up journaling for your entire organization because once you set up the policy, you still have to make sure that it is correctly applied on each of your servers. Once you do journal a database, all its mailboxes will be included. The other problem is that this solution captures only inter-organizational mail—only the mail that is going around the organization—but it will not necessarily journal things that are gatewayed out by SMTP. It will be captured on the sending system when the sender originates the message, *if* the sender's mailbox is on a database that is been journaled. But it is not necessarily going to give you 100 percent coverage, unless you also have journaling enabled on your bridgeheads. Microsoft has produced a document at <http://www.microsoft.com/technet/prodtechnol/exchange/guides/E2K3Journal> that describes what you can and cannot do with Exchange message journaling. We recommend that you look over this document to get a feel for how your Exchange implementation will influence what you do with your retention and compliance implementation.

Retention and Compliance Policy Pitfalls

Notice that, up until now, we have not mentioned searching, sorting, or timing out messages. Exchange does not perform any of these tasks. As a result, there are some relatively serious pitfalls that you have to be aware of.

Missing mail. By missing mail, we mean you might have mail that you do not catch that turns up to bite you later. If you miss mail at your retention kick-off, when you first start putting mail to be retained into your system, if you do not

capture all the mail that is sent or received, or if you do not retain mail that is captured by the journaling system but not retained, you might have a big problem.

Failure to deliver. Okay, we have all seen in television or movies, or read in books, the scene where the angry judge slams down his gavel and says, “I hold you in contempt of court, and you are going to jail for 10 days, or 30 days, or until you produce this information.” Well, the fact of the matter is, judges can do that, and sometimes they do, to organizations or people who cannot produce the requested records in a timely manner. What “timely” means will vary by jurisdiction and by the individual judge. But, in general, you never want to be in a situation where you cannot quickly produce the results you are looking for—even if all you are producing is just a report to show you what is available, so that you can then go back and say, “OK, here are the messages that we think are pertinent to this request. It will take us an additional 2 weeks to bring in all the backup tapes, restore everything, and get those messages. But at least we have some idea of the scope of what we’re talking about.”

Not being able to show a chain of custody for messages.

Your archiving system has got to be able to produce reports that show you when messages were captured, and what has happened to them since then, particularly if you are using a system that lets you purge messages. If a message was purged, you will want to see who purged it and when it was actually removed, just to make sure that you do not face any questions in the future about something being removed that should not have been.

Inconsistent retention policy application. Probably the biggest pitfall that most organizations face on an ongoing basis is spotty application of their retention policy. People who circumvent it, for example, by making local copies of messages that they keep on their computer in a PST file can create a big problem. People will make surreptitious copies of mail and file them away, just as sort of a self-coverage mechanism. There are organizational ways to deal with these problems, but from a technical standpoint, being able to control who can use PST files and what happens to them is extremely significant.

The other big problem is something that goes much more toward policy, and toward what we call an HR solution than a technical solution. If you have a policy, you have to make sure that it is consistently applied—to all the parts of your organization that it should apply to. Every business

unit, every subdivision of your organizational structure to which that policy should apply—you have to make a pretty strong effort to make sure that the policy is applied uniformly wherever it should be. Technical products, or technical solutions, will help with that; but more broadly, you also have to be prepared to do some old-fashioned HR work to make that happen consistently.

Best Efforts Compliance

For many IT professionals, you are not necessarily going to be facing the kinds of requirements that large companies who are publicly traded will face, or face the same requirements of those people who are working in industries that are sensitive by their nature—for example, government contracting, the healthcare industry, financial services, tobacco companies. So for many organizations, best-efforts compliance is “good enough.” Remember: we’re not giving you advice to tell you whether your organization falls into that category. What we *are* saying is that, for many organizations, it will be “good enough” to show that you have determined a policy, that you have notified everybody in the organization of what that policy is, and that you have taken reasonable technical measures to implement it, using third-party tools. To capture the message data in the first place, to use Exchange journaling to capture newly created messages that are originated or received after the kick-off date of your compliance and retention project, and then using the retention system to keep track of the messages on an ongoing basis.

Now, this sounds very simple, and we are glossing over some of the details that you would face in actual implementation, because those details will vary, depending on the nature of your implementation and the nature of your business. However, there are some first steps that you should consider taking as you explore what retention posture makes the most sense for your organization.

First, figure out what your policy is, or what it should be. This decision is going to be guided by the industry you are in and by the number of countries that you operate in (e.g., if you operate in the United States and the European Union, it is best to follow the most restrictive set of policies that are imposed on you by governmental regulations). The nature of your retention system will probably look different if you primarily do business in the United States, Canada, and Mexico, but determining your policy will depend on where you do business. What do your lawyers say about your policy? Do they know you have a policy? My experience has generally been that organizations that have retention policies have them because the lawyers woke up one

morning and said, “My Goodness! There are probably all kinds of stuff in our email system that we do not know about, or that we might not want to keep around.” What are other companies in your field doing? Now, normally we are not for arguing or talking about keeping up with the Joneses, but in this context, it is significant because what other companies in your field or in your industry are doing will help guide expectations about what is reasonable and customary in your industry. If everybody else who makes the same kind of widgets you do is keeping their mail for 3 years, and you are only keeping yours for 15 days, that looks unusual. It makes you look odd compared to your peer companies, and a judge or a smart attorney is perfectly able to raise the question of why you are doing things differently than your peers. Maybe it is because you are up to no good, or maybe it is because your organization does not have the technical competence that it needs to make sure that this is implemented properly. Of course, another broad question is what you actually can afford to do. We cannot answer that for you, but we think that it is reasonable to assess what it actually costs you to implement these solutions against the cost of not complying.

Next, once you have an idea of what your policy should look like, or what it does look like currently, you can start to pilot a retention solution. We always recommend that people pilot their retention solutions on their Exchange Administrators first. This gives you a captive audience, if you will, of people who know what the messaging system looks like, how it works, and what kinds of mail flow they *should* be seeing. If you look at your retention system, and you see “OK, I usually get about 100 messages a day in my mailbox, but for some reason, when I look at the retention statistics, I only see that I got 40 messages on average over the past month,” maybe that is indicative that something is not set up right. The other reason to do this is that you always want to practice on data that will not necessarily have an impact on your retention policy, longer term. So it is not a good idea to start with a pilot that involves capturing all the CEO’s mail until you are sure that your retention system and policy are working in harmony to capture the mail that you really want.

Next, make sure that the searching, reporting, and query features are doing what you want them to, and that they are producing the kind of data that you have to have the ability to produce. If not, that system’s not going to do you any good. And so you have to be very careful about it. Keep a good paper trail, showing what your policies are and how they were implemented, so that in the event you ever *have* to produce proof of policy, you will be able to. Once

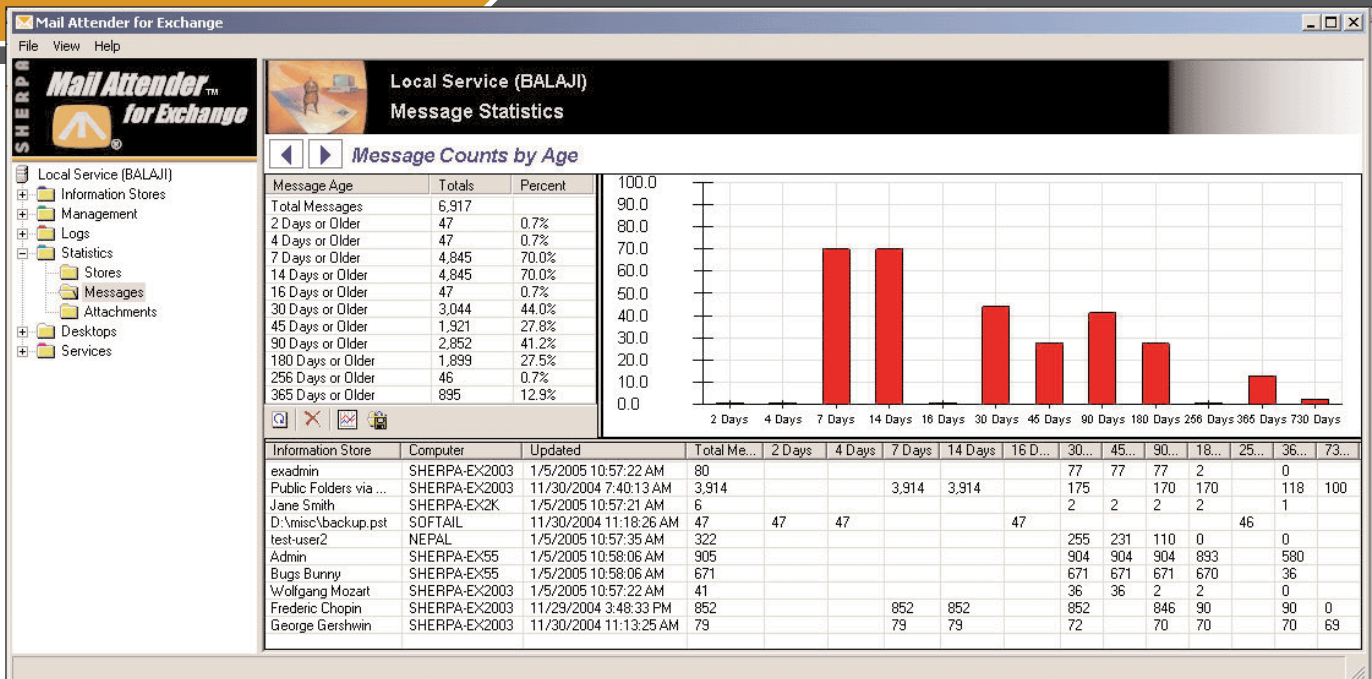


Figure 1: Sherpa Software Mail Attender

those things are done, then you are actually ready to enter into the deployment phase. Let us look at how Sherpa Software's solutions can help implement some of the requirements we have discussed to this point.

Addressing Email Storage Concerns

First, let us address some of the email storage problems facing companies today, and hopefully highlight some steps that you can take to solve these problems. We will begin by taking a look some results from surveys conducted by Osterman Research regarding email usage and storage. Despite all the buzz about email storage and archiving, only one in seven companies surveyed has implemented an archiving solution in their organization. That means the majority of companies are still doing discovery and email management the old-fashioned way. Many of you may have had to perform some of these discoveries within emails either for organizational or for legal needs, and you are probably aware of how expensive this process can be without an automated system to help you. In addition to the man hours required to perform these searches, you have probably read about the millions of dollars in fines some companies have paid for failure to conform to compliance requirements. So not having an automated solution in place can be a very expensive proposition. In addition, four out of five users, generally because of a lack of a better option within the company, create and manage private email archives. In the Exchange world, that usually means PST files.

So if you are trying to formulate a solution, where do you want to begin? Let us start by trying to determine how

bad the problem really is in your environment. Not knowing what is in your email can be quite dangerous. We have all read about the inappropriate and sometimes incriminating internal emails that were discovered and subsequently made public during some high-profile trials, such as Enron and Microsoft. Keeping tabs on your email can help you avoid being caught in some sticky situations like these companies found themselves in. Also, the flood of email, especially with the proliferation of spam and email with large attachments, such as music files and MPEGS, if left unmanaged, can and does lead to huge storage problems. Storing these types of files results in cost and performance issues such as the need for additional mail servers, delayed backups, and an overall inefficient email environment. And as the use of PST files becomes rampant in the company, the control an administrator has over email data in the organization reduces even more.

So how do you keep track of what is in your email system? A number of tools currently are available that can help you with this task. Sherpa Software offers one such product, called Mail Attender. Mail Attender can provide you with a high-level overview of all data stored in mailboxes, public folders, and local and network-based PST files. The data can be viewed and categorized by age, size, types of attachments, and many other options. And the information is also graphed, providing a more visual view of the data. This type of real-time snapshot can be invaluable if you are trying to diagnose issues with email storage and performance. A number of Sherpa Software's customers have used these statistics to get buy-in from upper management to take more permanent measures, such as deleting and archiving email. In addition to real-time information,

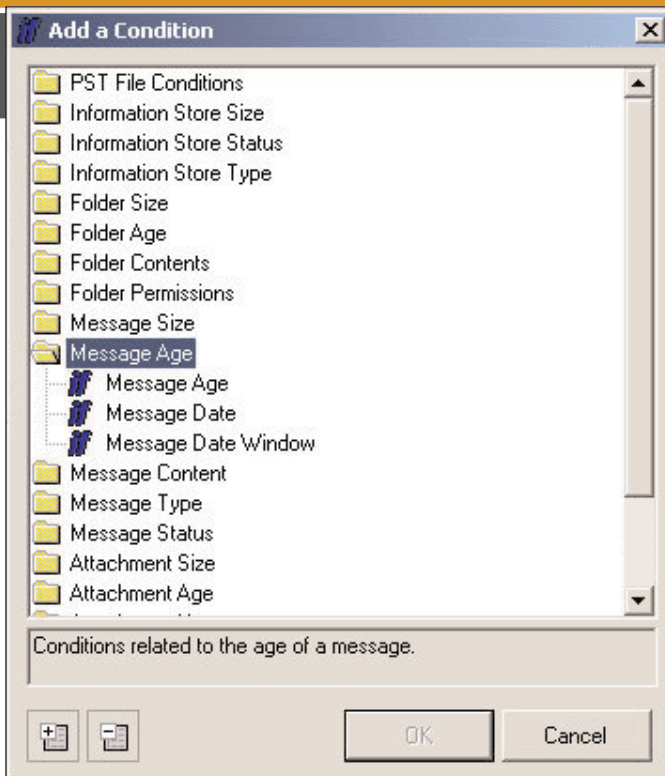


Figure 2: Using a message-age policy to delete expired email

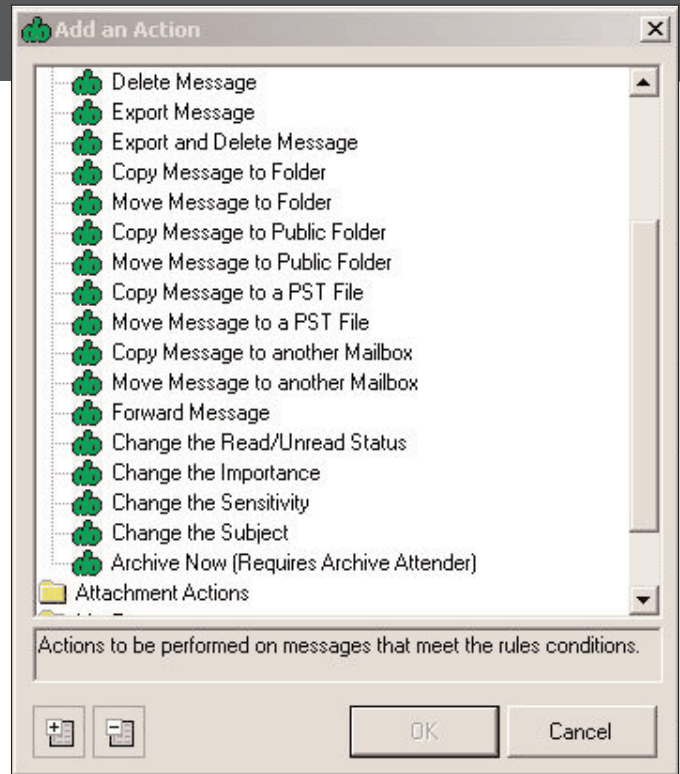


Figure 3: Configuring actions for a message-age policy

all this data can be gathered over time to help track trends in the growth of the email store within the company.

Mail Storage Best Practices

We have talked about the importance of knowing what is in your email storage. In this section, we will discuss the steps you can take to manage your mail, such as setting up retention policies to cleanse your system of all unnecessary email data. In all the years that Sherpa has been developing products for email management, we typically see companies try to address the email storage problem inhouse. The most obvious and politically safe solution is to add email servers. Although this option might be viable in the short term, most companies realize the need for something more permanent. Implementing a companywide policy of restricting mailbox sizes, generally referred to as quotas, is theoretically a very good option, but it generally requires a lot of political wrangling and, depending on your corporate culture, may not be something that you might be able to implement. Enforcing quotas has another undesired side effect: Users begin to generate and use local PST archives to store overflow email from their mailboxes. Another common solution we have seen customers try is to use built-in features of Exchange. Although these solutions may work for some situations, the options available are limited and rudimentary. Most companies typically need a more robust, automated solution that has an array of flexible options to help clean out their system of unwanted data. Sherpa's Mail Attender is one such product, as Figure 1 shows.

Sherpa Mail Attender

Using Mail Attender, you can search and discover unwanted emails, then take appropriate actions against them. Mail Attender lets you narrow your searches by age of messages, keywords within the messages and attachments, types of attachments, message and attachment sizes, and many other filtering options, giving you very granular access to all email data. As for actions that can be taken, in addition to deleting messages and attachments, Mail Attender along with Archive Attender provide several reporting, copying, moving, and archiving options. To give you a real-life example of the use of Mail Attender, one of our customers has a policy in place in their company to delete all email older than 6 months from all their employees' mailboxes and email older than 1 year from all their executives' mailboxes. For all users that have email that needs to be kept in the mailbox, the policies are automated to skip a folder named Do Not Delete. As Figures 2 and 3 show, you can see some of the options to execute this message-age policy to delete these email messages.

Mail Attender can target searches to specific Exchange mailboxes, public folders, and desktop and network-accessible PST files. Searches can also be performed on Exchange distribution lists. In the case of the customer we mentioned above, they use Mail Attender to target a policy specific to all the users, then use a separate policy for all their executives.

You can set up a policy in Mail Attender to run against all mailboxes found on a particular Exchange server or a

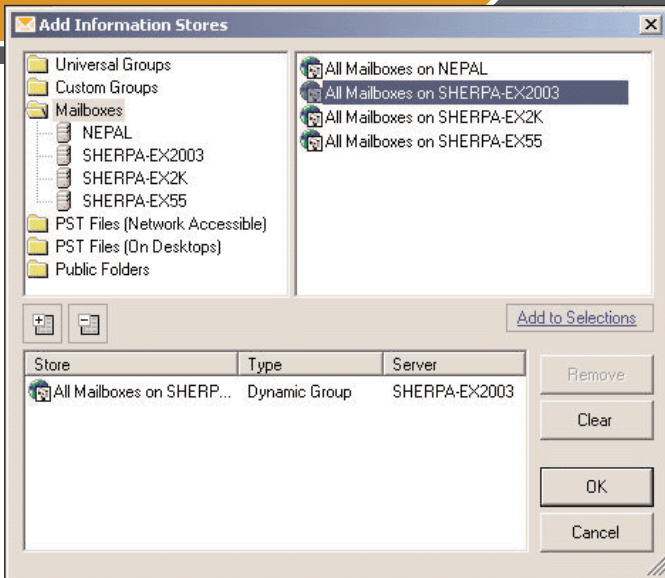


Figure 4: Configuring scheduling options

subset of mailboxes. Setting a universal group, such as All Mailboxes, ensures that, if mailboxes are added to or removed from the Exchange server, the policy will be automatically updated and executed on the appropriate mailboxes by Mail Attender.

You can automate the Mail Attender policy to run at the most convenient times in your organization. Figure 4 displays the various scheduling options that are available.

In using these options, you can enforce policies at times that least affect the performance of the mail servers. In this example, the policy will be enforced once every hour, every

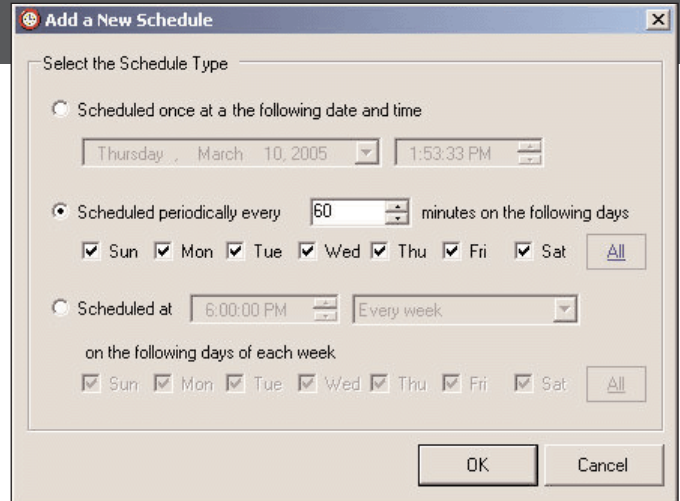


Figure 5: Configuring timed policies

day of the week, as Figure 5 shows.

In the case of the customer we have been talking about, they have it set up to run at 8:00 p.m. every evening. You can see the importance of using a solution to automate some of the process of cleaning up your company's email stores.

Tackling PST Files

And now a little bit about the other management headache for IT administrators, dreaded PST files. As we have already discussed, users generally create a PST file to manage personal archives of their email, usually to offset quota limits on their mailboxes, or because of a lack of a better archiving option in their organization. Just by their very

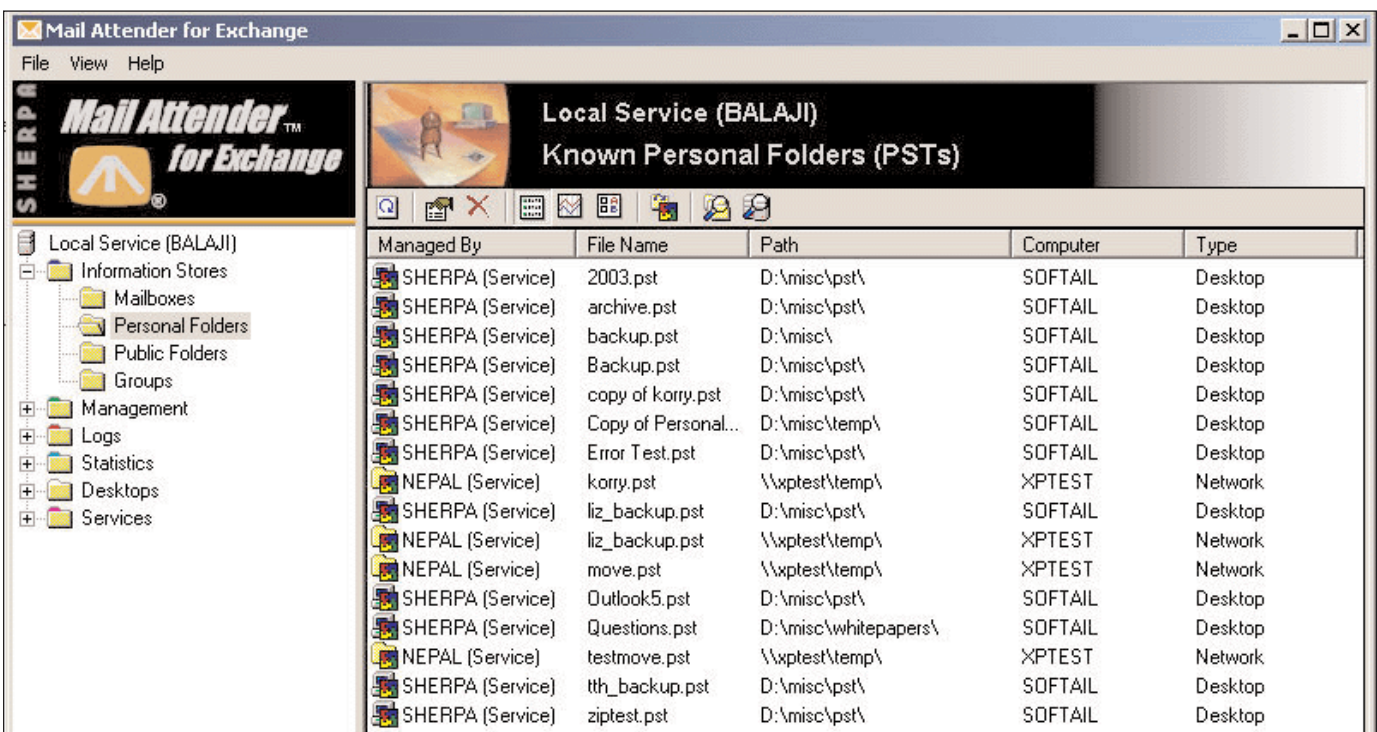


Figure 6: Gathering .PST file information

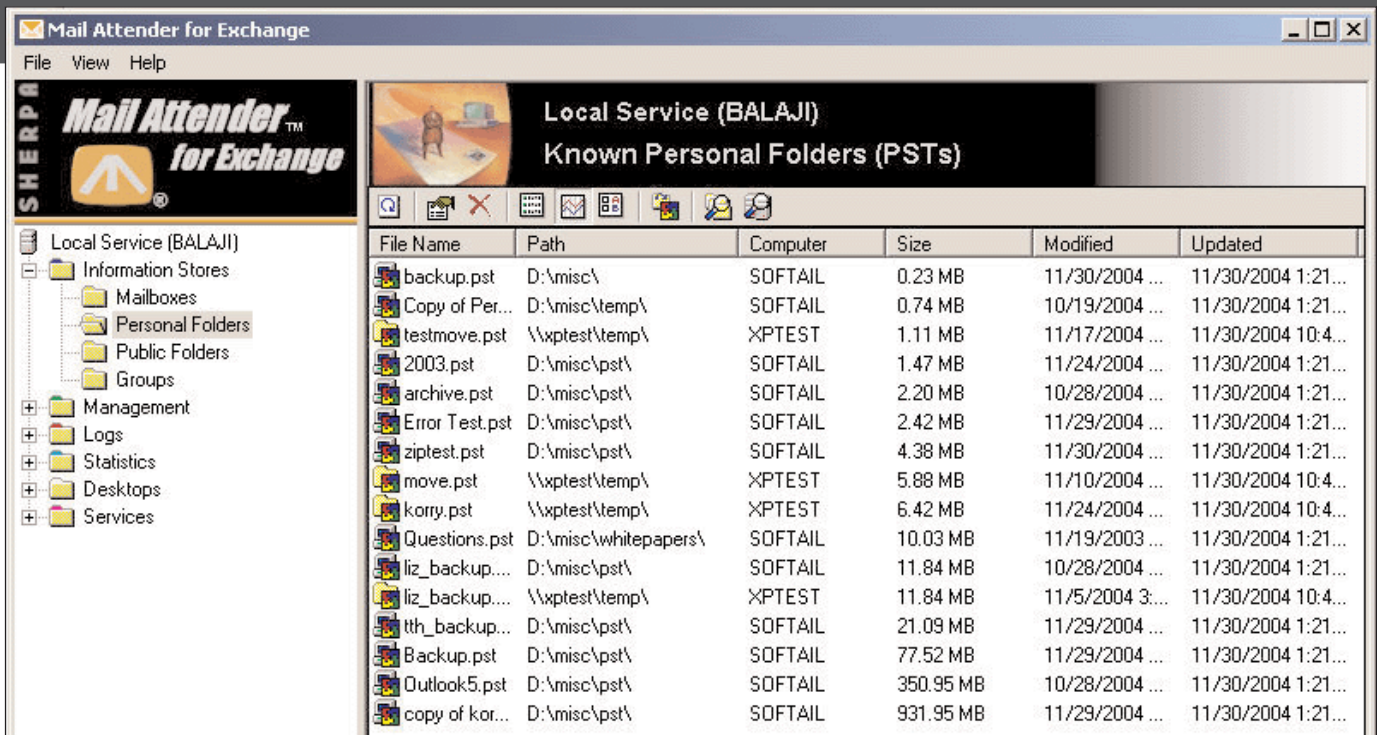


Figure 7: Gathering additional PST file information

nature, PST files are decentralized, making them hard for administrators to find, and even harder to search and manage, especially if they are stored and maintained locally on users' desktops. Also, PST files created with earlier versions of Outlook have had a history of problems with corruption leading to loss of data. We have actually had numerous customers who had terabytes worth of PST files taking up space on their files servers, and with no easy way to look through them, had no way of going through and cleaning out the unwanted data to recover any space that the files were using. In fact, greater than 60 percent of all Sherpa prospects come looking for a solution to either manage or eliminate PST files from their environment. So if you are grappling with a PST file issue, rest assured you are not alone. There are products available that can help you get a handle on this problem.

Mail Attender offers several features to manage PST files. As Figures 6 and 7 show, Mail Attender displays the types of information that you can gather about PST files, both local and network based.

As new .PST files are created in known locations, Mail Attender will automatically include them in its list of files to monitor, providing the ability to manage them and take control over them.

With Mail Attender, you can set policies based on size and access dates of PST files, in addition to age of messages, size of messages and attachments, specific keywords within the messages, and many other criteria. Mail Attender lets you delete or export content, and can automatically compact the PST file to reclaim the recovered

space immediately, as Figures 8 and 9 show. In addition, should your environment not allow deletion of PST files, Archive Attender can archive from PST files.

One of the customers that we referred to earlier had more than 2TB of PST files and was able to use Mail Attender to delete all the messages in those PST files and reduce the amount of space required by these files to less than 600GB. So do not despair: If you are one of those stricken by PST-itis, there is hope.

Just to recap, we have talked about the importance of knowing what is in your email stores, using either inhouse policies or third-party tools to purge your system of unwanted email, and getting a handle on all the PST files in your environment. The next step is to decide what to do with the information you need to keep.

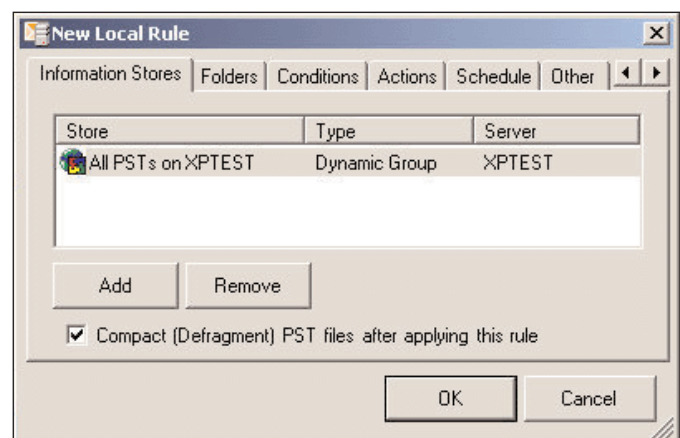


Figure 8: Compacting PST files to reclaim recovered space

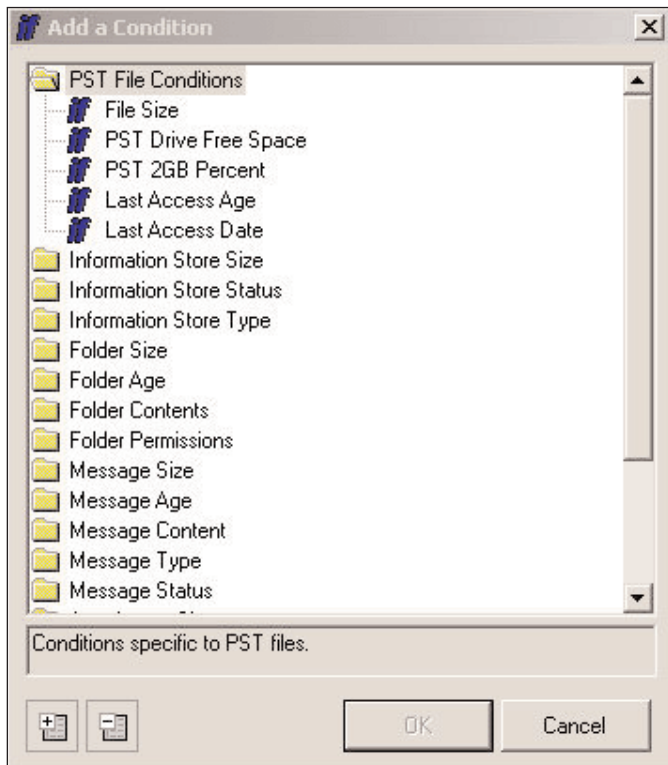


Figure 9: Configuring conditions for the compaction

Managing the Data You Keep

As we have already touched on, Exchange has very limited options for archiving email. We have already talked about using PST files as an archive, and the inherent problems with them. You can archive messages to public folders, but this does not do anything to save you space on the email servers. Another option that Exchange offers is journaling, which we've already discussed. Many of the archiving solutions available in the marketplace today use journaling. However, without the ability to actually examine the contents of the journaled email and act on it, journaling in and of itself just compounds the email storage problems.

So given the fact that Exchange does not give you a whole lot of options, your best bet is to look for an archiving solution that best meets your specific requirements. One of the first questions you need to answer before you begin looking for a solution is, "Are you archiving to recover storage, are you archiving because you are required to comply with industry and corporate regula-

tions, or both?" We will look at the features of each of these separately, beginning with storage archiving.

Storage Archiving

Generally, the reasons companies archive for storage are self-evident—to recover space on the mail servers and also to extend the size limitations of the Exchange server. With more and more users using their mailboxes as filing cabinets, providing them with an archiving option redirects the storage to an external device, relieving the Exchange server of the excess data. Because storage archiving directly impacts the user, it is imperative that the solution you select provides a way for users to easily view and search the archived messages—if possible, from within the existing Outlook interface. It should also include the ability for them to archive their own messages, hopefully eliminating the need for them to use PST files. In today's highly mobile environment, if you have traveling users, access to the archives through Outlook Web Access (OWA) should be another requirement in your archiving solution.

Sherpa's Archive Attender offers a solution for companies looking to reclaim space on their mail servers, giving users access to their archived messages and the ability to search their archives from within Outlook, without requiring any client installation. Figure 10 shows Archive Attender's view of a user's archived messages, as seen through his or her Outlook interface.

Archive Attender also provides mobile users access to their archived messages when they are connected using OWA.

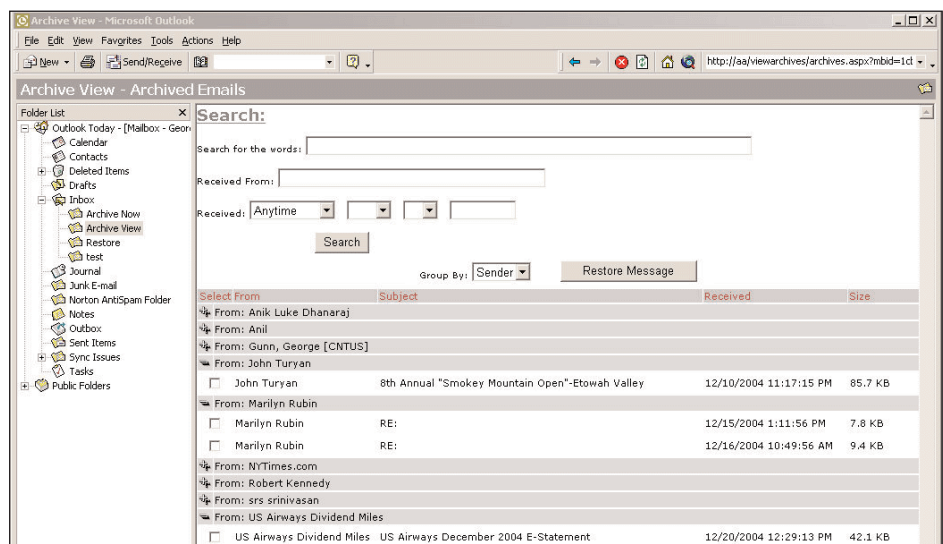


Figure 10: Viewing archived messages

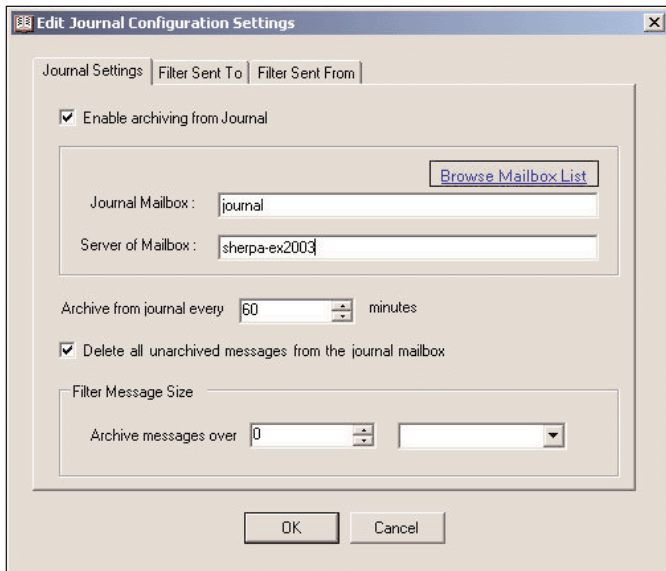


Figure 11: Configuring journal settings

Compliance Archiving

The other reason companies archive data is primarily because they are required to store information and keep it secure and intact in its original form for extended periods of time. Although regulations vary, and most of them are not very clear on exactly what information needs to be retained, there are several general features that must be part of any compliance-based archiving solution, including the ability to target specific senders and recipients, and specific keywords within email that are to be archived. Once archived, the information must be maintained securely, be indexed to facilitate quick and efficient searching, and also maintain an audit trail of all activity within the archives.

Figures 11 and 12 show Archive Attender's ability to take advantage of Exchange's journaling, to capture and archive either all or a subset of all incoming and outgoing messages.

If not all messages are archived, those left behind in the journal will automatically be deleted by Archive Attender, thus preventing exponential growth in the size of the journal mailbox.

Targeting Content to Be Archived

Regardless of the reasons you are archiving, you need to make sure the solution you choose gives you a great deal of flexibility in filtering content within the mailboxes by letting you select different properties to search on, and also to use multiple search criteria to make the discoveries more accurate and efficient. This capability is especially necessary in the case of compliance archiving, because, as we

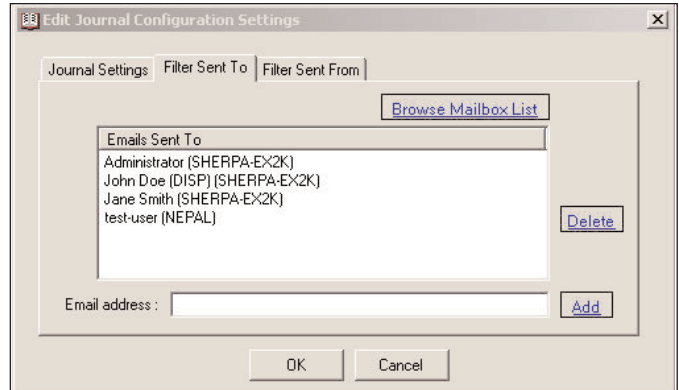


Figure 12: Viewing journal filter settings

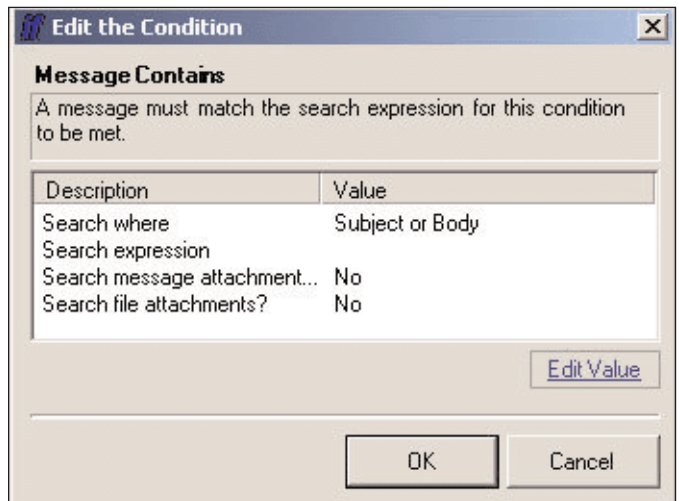


Figure 13: Editing the search condition

mentioned before, many of the regulations are not very clear around what information you need to keep. Figures 13 and 14 show the flexibility of Archive Attender's search capabilities. In addition, Archive Attender and Mail Attender are integrated to include extensive content selection options, such as by keywords, attachment type/size and other customizable criteria.

One of Sherpa's customers, a large energy company on the West Coast, is governed by the Federal Energy Regulation Commission. They must repeatedly search their email stores to find additional information each time their compliance guidelines were being audited. They have since purchased Mail Attender from Sherpa and have greatly simplified this discovery process.

The Information Life Cycle

After the content has been moved into the archive, the final step of managing the content entails processes generally referred to as *information life-cycle management* and

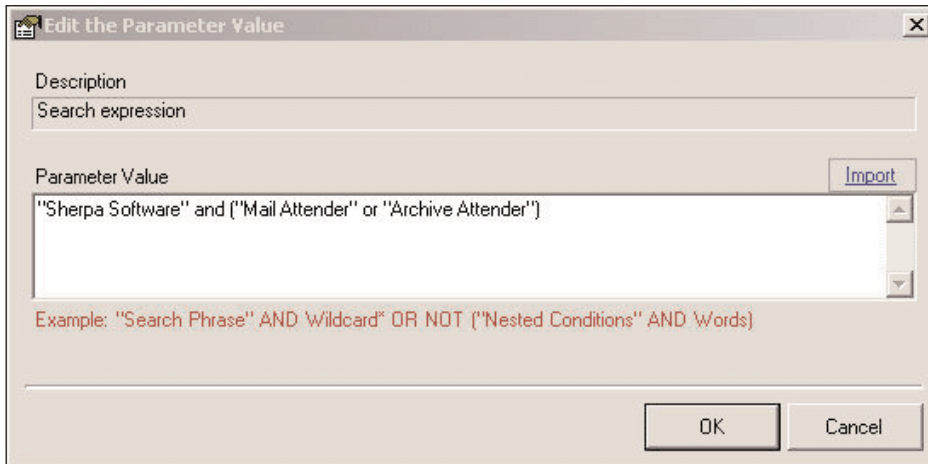


Figure 14: Configuring the search expression

hierarchical storage management. In addition to discovering and retrieving content within the archive, you need to decide for how long you want to keep the data within the archives, and the action to be taken once these retention periods have expired. Depending on the needs of your organization, and the industry regulations governing you, you might be able to delete some of the older content and/or move other files to offline storage. For instance, financial documents that meet certain criteria are required by the FCC to be retained for a period of 7 years. HIPAA regulations require that certain medical documents be kept until a patient reaches 21 years of age. Securing and encrypting the data within the archive is another important task that might be dictated by the regulations. These are additional steps to consider when you are looking for an archiving solution to implement in your company.

Sherpa Solutions for Email Management

Mail Attender has been Sherpa's flagship email content-management product that is administrator driven and can be used to report on and clean out existing mailboxes, PST files, and public folders by targeting content that meets specific administrative criteria. Archive Attender is a data-archiving solution that provides features to archive email to network storage devices for storage or compliance; to maintain an index of all stored email for efficient searching

of the archives; and the ability for users to access, search, and restore their archived messages. These two products function independently or can be integrated to provide an automated, robust, flexible, and complete email-management package that can be used for everything from finding email, to deleting unwanted email and archiving email that must be kept.

Free, fully functional evaluation versions of both products are available on our Web site and can be downloaded and deployed in less than 30 minutes. As we've men-

tioned, Mail Attender provides a comprehensive solution to manage PST files, typically a scourge for most IT administrators. Unlike many solutions available, Mail Attender and Archive Attender require no components to be installed on the Exchange Server, nor do they require any additional hardware. The products are priced moderately, starting at \$14 per user, with decreased pricing for purchasing higher quantities.

Please feel free to visit our Web site <http://www.sherpasoftware.com> for information about us and our products. ■

About the Authors

Paul Robichaux is a principal engineer for 3sharp, an MCSE, and an Exchange MVP. He is the author of several books, including *Secure Messaging with Microsoft Exchange Server 2003* (Microsoft Press), and creator of the <http://www.exchangefaq.org> Web site.

Balaji Srinivasan is a Senior Product Manager for Sherpa Software's Exchange group. His is responsible for designing, building, testing and documenting new products, as well as enhancing existing products for the Microsoft Exchange platform. Bal has extensive experience in product development, having worked for over eight years as a Senior Product Developer in the consulting industry.