12 Minute Guide to Email Archival Search
Foreword

Too many whitepapers spend too much time building up to the meat of the matter. With that in mind, X1 created this concise guide to address a specific topic, that of how to find email and associated data once you implement an email archival solution.

While email archival solutions are of great benefit when it comes to managing the growth and meeting compliance regulations around data retention, they can complicate the process of finding data that exists in primary email systems along with online and offline vaults.

Contents

Introduction ................................................................. 1
Why Email Archiving Exists in the First Place ................................. 1
Enterprise Search Federation Delivers Results ................................. 2
A Tale of the Email from 2003 ............................................. 4
Sponsored by ................................................................. 6
Developed with ............................................................... 6

Figure 1: Federated search ................................................... 3
Figure 2: X1 Enterprise search architecture ................................... 5

©2008 Grey Consulting and X1 Technologies, Inc. All rights reserved.
Introduction

Increasingly, organizations are implementing email archiving to bring order to email messages. In tandem, federated search, a term that is being talked about by leading industry analyst groups such as IDC, is gaining ground as a means to provide a unified view of search results within an enterprise. Federated search enables information access and retrieval for archived email and other types of data in disparate locations. Email archiving, along with federated search, provides a viable solution that addresses the needs of the knowledge worker, IT and corporate governance.

Effective management of email information requires an understanding of:

- Internal email policies combined with regulatory governance
- How email archiving can control the growth of primary email applications and address corporate retention policies
- How federated enterprise search can provide a means to find data

Why Email Archiving Exists in the First Place

Email archiving solutions capture and store a pristine copy of inbound and outbound email messages and attachments. Messages are indexed, and a server-side search engine allows the user or administrator to search for and retrieve specific email messages through a Web client or, more commonly, a desktop email client, e.g., Outlook.

To the end user, where the email exists is immaterial. Most enterprise email archival solutions, such as Symantec Enterprise Vault make the process seamless by providing a “stub” for the archived data. Even though the data has been moved to a “vault” for archival, the end user still sees the data as residing where it always has.

Due to a stringent regulatory environment, U.S. financial service providers were early adopters of first-generation email archiving products and services (1996 to 2001). By the mid-2000s, email messages were often called into evidence in regulatory, legislative, civil and business proceedings – across market sectors.

These regulations along with the litigious nature of many businesses have caused enterprises to set conservative policies for retention, assuming the best policy is to have a “better safe than sorry” policy.

Declaring an email message as a record is starting to become common practice, and, as such, email information is considered to be a critical component of the business infrastructure.

Another distinct driver for email archival is the need to control the growth of email on primary email applications such as Microsoft Exchange Server and IBM’s Lotus Notes. Many IT organizations have struggled to manage their email systems due to annual growth rates of 50% or more. The larger an email system becomes, the harder it is to perform critical functions such as backup and recovery.
Given that many IT organizations are cost centers within a business, they are held to high expectations of end users who don’t want to be constrained by mailbox size quotas and message sizes, yet don’t want to be burdened with the costs of extra storage and server.

In summary, email archiving can address multiple issues:

- Address local and federal government regulations
- Meet internal service levels for data retention and management
- Increase end user productivity by removing mailbox quotas
- Improve backup and recovery operations of email systems

**Enterprise Search Federation Delivers Results**

Business search is led by awareness. The individual may remember creating or reading the desired information. The likelihood is that metadata (data about the data) will be used to help define the search criteria. This is quite different from a generic search, such as conducted on the Internet, in which the individual uses a variety of keywords led by the hope that the desired information exists.

Enterprise Search solutions with federation circumvent data disparity.

However, this awareness (“I know it’s here somewhere”) can lead to frustration. Information-generated through email or other utilities-can be created by one or multiple applications and be stored in one or multiple data repositories.

Information management nirvana would be one huge, converged data repository populated by information created by every application used across every business unit--not a reasonable quest given the sheer quantity of information being created every moment and the growth of applications that create the information.

With the growth and maturation of Enterprise Search, a market that is growing 20-30% annually and will grow from $1.7 billion in 2007 to over $3 billion in 2012 according to IDC, businesses need not wait for nirvana. At the core of a true enterprise search solution is a federated (unified) search engine that can circumvent the data disparity issue. Federated search technologies use “connectors” for information search and retrieval to multiple proprietary data repositories such as email, databases, email archival systems and unstructured data on PCs and servers.

A federated search combines multiple searches from multiple repositories into a unified view.
The majority of applications include a native search engine, which is used to find information in that application’s data repository—five applications equal five native search engines. Federated search uses multiple engines across multiple application data repositories and consolidates the results into a single view. Federated search results in:

- Greater likelihood of detecting related information
- Faster overall search time
- “Single pane of glass” user experience

In addition, native search applications are not all the same. Some have cumbersome interfaces that require extensive end user training while others are slow. And most only provide the ability to search within a specific application versus the enterprise as a whole.

This has driven the explosive growth of the Enterprise Search market which is predicted to grow to over $2 billion in revenue in 2008\(^1\). While Federated search is key to solving the problem of finding data that is both archived and still residing within the primary application, it is one of many features that allow end users, or “knowledge workers” a portal into the data stored throughout an enterprise.

Along with search federation, Enterprise Search solutions generally provide the following:

- Ability to search on numerous data fields including meta data attributes
- Sort results based on one or more data attributes
- Preview results in the native application’s format
- Support existing security policies
- Support a variety of data sources including structured and unstructured data
- Provide a “holistic view” of all data by searching both network attached servers and PCs
- Low overhead from both an index storage and PC/Server perspective
- Capability to search, preview and act upon data from a single window
- Allow for the execution of commands and actions native to a specific application
- Federate search results not only from multiple applications, but multiple sources such as PC, Server and indexes of other search engines

---

1 Source, IDC Worldwide Search and Discovery Software 2008-2012 Forecast
While many people still think of search in the terms of internet search, search behind the firewall is much more complex, and much more important to businesses trying to find information spread throughout the enterprise.

**A Tale of the Email from 2003**

What happens when you begin archiving email and out of the blue an urgent request comes from your legal department that they need all emails and files associated with a company called ACME INC, including attachments, from as far back as 2003? Since you are the account manager for ACME INC, this doesn’t sound like too hard of a task – albeit time consuming.

At first, you may simply go into Windows Desktop Search in order to find emails for ACME INC in Microsoft Outlook and those on your PC. Then you jump over to your archive solution, Symantec Enterprise Vault, to find the applicable archived emails. However, a couple of things become apparent. You discover that you need to know which Vault(s) the data for ACME INC reside in. You realize that you need to search not only your PC and associated PST files, but the shared file servers and data stored within Microsoft SharePoint.

Now you have to find emails for ACME INC that have been archived in Symantec Enterprise Vault, any current emails that exist in Microsoft Exchange, any documents that may exist on the company’s shared file servers and any stored within Microsoft SharePoint – and these are the repositories you know about.

**Without a federated search solution:**

You will end up searching dozens of data repositories on multiple servers, in local and remote locations and with no means to assure you found everything. You could utilize point products such as the native search within Symantec Enterprise Vault along with Windows Desktop Search and Microsoft Search Server 2008, but the method and results would be inconsistent.

Each search interface is different and the results will have to be aggregated manually. The time and effort to find all the data associated with ACME INC within your company has now become a major, time consuming project. In fact, the Association for Information and Image Management (AIIM) calculated that where the cost to buy one 1GB of disk storage is only 20 cents, the cost to review 1GB of storage is around $3,500.

**The solution:**

Solving this problem, which is becoming more common with each passing day, is the X1 Enterprise Search Suite that delivers federated search results via a single user interface. The Suite consists of the X1 Enterprise Server that consolidates search results from multiple locations via Content Connectors, providing Federated Search results. X1 has developed a broad set of Content Connectors to find data in popular email applications such as Microsoft Exchange and Outlook along with content management and email archival products such as Microsoft SharePoint and Symantec Enterprise Vault.

Lastly, the award winning X1 Professional Client delivers an intuitive interface that simplifies the search process with a unified view that shows the results in a unified view with the ability to both preview and act upon the search results.
The figure below shows an example of a federated search encompassing current email along with archived email (both offline and online) within Symantec Enterprise Vault, file servers and Microsoft SharePoint.
X1 Technologies was founded in 2003 to help business users easily find, preview and act upon information residing anywhere on the desktop or across the enterprise. X1’s search solution allows workers to access all forms of corporate data, from email in inboxes to contracts in document management systems to information in CRM and ERP systems, all through a single, elegant interface.

The core of the X1 Technologies solution, which is protected by one patent\(^2\) along with five pending patents, uses a proprietary method for indexing the content of every email, attachment and file on a business user’s hard drive, including network drives that appear on the desktop. The final index, which updates automatically based on each user’s settings, is stored dynamically. This lets X1 find, list, and display results rapidly, so users can find information they know exists without having to know how or where it is stored.

X1 Technologies is an operating company of Idealab and is headquartered in Pasadena, CA.

Grey Consulting is an independent research, advisory and consulting firm in the electronic communications, collaboration and social media markets. Topical coverage areas include product evaluation, industry compliance, data security and marketing. With over 20 years of analytic and operations experience, we work with our clients to help them make smart decisions.

\(^2\) U.S. Patent No. 7,370,035