Whitepaper: Journaling: There’s More Than One Way to Retain Your Data

September 2012

Prepared by Dan Levine, Principal Engineer
Globanet
15233 Ventura Blvd., Ste. 1140
Sherman Oaks, CA 91403
888.427.5505 | www.globanet.com
# Table of Contents

- Introduction ..................................................................................................................... 03
- Enterprise Vault Journaling ............................................................................................. 03
- Enterprise Vault Custom Filtering: Selective Journaling ................................................... 04
- Enterprise Vault Custom Filtering: Other Uses for Basic Filters ....................................... 04
- Enterprise Vault Advanced Custom Filtering: Compiled Filters ......................................... 04
- Enterprise Vault Data Classification Services ................................................................... 05
- Key Takeaway .................................................................................................................. 05
- About Globanet ................................................................................................................ 05
Introduction
The debate over whether organizations need to retain their electronic data in order to adhere to regulatory and other guidelines has been settled for quite some time. However, the methods and technology used to enforce retention policies are frequently subject to discussion and interpretation. A substantial percentage of organizations subject to retention obligations elect (or are mandated) to journal their electronic mail. A common misconception is that all of this mail must be journaled in the same way. Globanet can work with your organization to better understand what your retention obligations are and help you formulate a journaling strategy to satisfy them.

The most basic means of journaling occurs independent of any archiving software. All messages sent and received by members of a messaging system, such as Domino and Exchange, can be duplicated and kept locally on the email server, but separate from end-user mailboxes. Isolating a so-called “golden copy” of every message prevents users from deleting critical or potentially damaging information. That data can then be backed up and/or stored in offsite disaster recovery facilities. Unfortunately, this primitive methodology has intrinsic flaws that render it effectively useless. Some of these are listed below.

• **Mail servers are not file servers:** Messaging systems were not designed for long-term storage of data and/or document management; they are built for short-term storage and review followed by deletion.

• **Email solutions are not archiving solutions:** While they can retain data in the strictest sense, email systems aren’t built to do it well. This ranges from lack of de-duplication functionality to the inability to apply different retentions to different types of data.

• **Fast disk is not cheap disk:** Mail servers frequently use fast, expensive disks with high input/output loads, allowing huge quantities of data (in duplicate!) to accumulate, which is costly and wasteful. It also leads to decreases in server performance.

• **eDiscovery is hard work:** Searching through hundreds of gigabytes to terabytes (and more) of email is an intense task, particularly if the server hosting that data is actively serving mail. Virtually no major email application has any truly effective means of natively performing ediscovery. In the meantime, organizations are subject to increasing obligation thanks to a widening array of internal, regulatory and legal requirements.

Enterprise Vault Journaling
It is reasonable to dismiss on-email-server journaling for the previously mentioned reasons, but its general principals should not go overlooked. Journal mailboxes are the foundation for most respected journaling solutions. On a basic level, Symantec Enterprise Vault journal archiving uses the same journal mailbox as the source of its data. By removing data from the email server after it is ingested, Enterprise Vault journal archiving eliminates many of the issues discussed above. For example:

• **Enterprise Vault servers are not mail servers:** Mail is extracted from the journal mailbox and archived by the Enterprise Vault server in a different, usually less-expensive storage solution.

• **Enterprise Vault is an archiving solution:** Enterprise Vault is built to single-instance data and efficiently apply retentions in accordance with organizations’ legal and regulatory obligations.

• **Enterprise Vault is storage-agnostic:** Enterprise Vault can leverage virtually any storage solution, from CIFS shares and NTFS volumes to WORM devices, like NetApp Snaplock and Centera devices. Storage is only as expensive as an organization’s legal obligations demand. A small business seeking to retain data purely for human resources or small ediscovery actions can use SAN disk, file shares or even local drives, while a financial firm subject to stringent SEC and FINRA requirements might need more expensive WORM storage for regulatory compliance.

• **Enterprise Vault does eDiscovery:** Between Enterprise Vault’s basic Discovery Accelerator search module and Clearwell, a comprehensive, end-to-end legal review module, Symantec provides a range of mechanisms for performing crucial ediscovery functions with no impact to active Exchange users’ mail.
Enterprise Vault Custom Filtering: Selective Journaling
Traditional Enterprise Vault journal archiving is simple: All mail sent or received is archived for a customizable retention period from zero days to 999,999 years (“Forever”). It is indexed for improved searchability and de-duplicated. For many customers, particularly smaller ones, this approach is sufficient. Other customers have more specific needs, which Globanet can address in a variety of ways. One such tool is selective journaling, which allows the journal archive task to parse messages based on a variety of criteria and apply different retention categories (and more) accordingly.

Selective journaling is as close as the Enterprise Vault journaling process comes to duplicating provisioning, a component of another Enterprise Vault module known as mailbox archiving. Journal archiving targets just the journal mailbox, while mailbox archiving targets end-user mailboxes, using provisioning to granularly apply retentions to different users and/or groups. For example, an Enterprise Vault administrator might choose to retain the contents of executives’ mailboxes for 10 years, while managers’ items are retained for five years and average users’ messages are retained for only one year. It is important to understand that customers who employ mailbox archiving frequently expend substantial time and effort when architecting these groups and frequently employ journaling as well.

There is no direct analog to provisioning in journal archiving, though. Multiple journal mailboxes can be created by targeting specific mail databases or servers, but in most situations, all messages sent or received by the users represented by a journal mailbox (and thus, journal archive task) are treated identically. This may suffice for many customers, but others require greater specificity or – in the case of an organization using both mailbox and journal archiving – consistency of retention across the Enterprise Vault platform. That is, a customer’s legal department may require that all messages received by executives are archived under the same retention category regardless of whether they were archived by the mailbox archiving module or the journal archiving module.

The need to address different users and user groups differently can, of course, develop even if mailbox archiving is not present in the environment. Globanet engineers can assist customers in defining and writing the custom filters necessary to provide granular retention through selective journaling.

Enterprise Vault Custom Filtering: Other Uses for Basic Filters
Selective journaling is only one way that custom filtering enhances journal archiving. Customers may use custom filtering to perform a variety of actions, including redirecting certain types of messages (such as meeting acceptances) to different archives or excluding other types of messages, such as voice messages, system messages and read receipts from the archiving process. Filters can also insert a variety of index properties that can be acted upon later, such as during ediscovery searches. These basic custom filters are typically smaller undertakings and can be generated by Globanet with a relatively lower level of effort.

Enterprise Vault Advanced Custom Filtering: Compiled Filters
Globanet can also leverage our development team’s substantial expertise in generating compiled filters, which can perform significantly more complex functions. Like basic filters, these filters also “watch” the journal mailbox and trigger events; the difference is that compiled filters can make API calls to other applications, not just actions within Enterprise Vault. One such filter is Globanet Identify, which acts on image files. Those files are sent to an optical character recognition (OCR) application, enabling Enterprise Vault to index images. This filter independently provides oversight into large quantities of data that would otherwise be excluded from indexing. Globanet Identify is just one example. Our team can develop compiled custom filters that integrate Enterprise Vault with other applications, databases, document management systems and even different archiving systems.
Enterprise Vault Data Classification Services
Globanet Custom Filters provide our customers with opportunities to expand Enterprise Vault’s native journal archiving functionality at a reasonable cost with no additional hardware requirements. Of course some customers have journal archiving requirements that require a depth and breadth beyond the scope of a typical custom filtering solution. In these instances, Globanet recommends Enterprise Vault Data Classification Services, a module that yields results that are ultimately similar to those provided by custom filtering but with increased granularity and customizability.

Data Classification Services’ rule-based interface enables organizations to perform a variety of actions based on the metadata of messages it receives. The Data Classification Services application comes with a variety of built-in rules, including the ability to search for and auto-tag Social Security or credit card numbers sent via email as well as the capacity to create custom rules (and modify existing ones). This flexibility puts more power in the hands of Enterprise Vault administrators. Messages can be automatically deleted, ignored by the archiving process, stamped with a designated retention period depending on their content or otherwise tagged pursuant to the defined rule. Unlike custom filtering, however, Data Classification Services requires additional hardware, licensing and implementation considerations, which can be more expensive.

Key Takeaway
Enterprise Vault journal archiving provides an excellent framework for archiving journaled email messages, but unsurprisingly, there is no one-size-fits-all approach. No matter what your organization’s requirements are, Globanet’s industry-leading expertise enables us to provide a solution tailored to fit your unique needs.

About Globanet
Globanet is a leading provider of archiving and ediscovery solutions worldwide. Founded in 1996, the company is a pioneer in archiving and intelligent information governance, and has developed a portfolio of software and services to help organizations manage data from creation to expiry. Globanet’s proprietary solutions include the Merge1 message capture platform and Migrate (formerly GEM) data migration software. A Symantec Platinum Partner with Master Specialization in Archiving and eDiscovery, its professional services team has extensive experience with industry-leading Enterprise Vault and Clearwell. Globanet’s broad range of services includes policy and solution design, installation and configuration, data migration, custom add-ons and project-based ediscovery consulting. For more information, call 888.427.5505 or visit www.globanet.com. Visit the Globanet Blog at http://www.globanet.com/blog or follow us on Twitter at http://twitter.com/globanet.