

Efficient and Cost-effective Email Management With XML

Dr. Johannes C. Scholtes, President and CEO, ZylAB North America LLC



Dr. Johannes C. Scholtes

Dr. Johannes C. Scholtes is the president and CEO of ZylAB North America LLC and is in charge of ZylAB's global operations. Since Scholtes took over the leadership in 2002, ZylAB has enjoyed double-digit expansion as well as consistent annual growth in profitability. Before joining ZylAB in 1989, Scholtes was an officer in the intelligence department of the Royal Dutch Navy. Scholtes holds an M.S. degree in Computer Science from Delft University of Technology (NL) and a Ph.D. in Computational Linguistics from the University of Amsterdam (NL).

Complying with mandated e-discovery requirements necessitates the use of XML platforms for optimal email management and storage. Multiple factors exist regarding how organizations approach archiving email in terms of their overall email management plans. Typically, organizations must consider whether they regard email storage as fitting into one or more of the following contexts:

- ◆ An IT/storage issue—focus is on saving disk space;
- ◆ A customer service/sales issue—focus is on providing better customer service by having all customer communication on-line;
- ◆ A project management issue—focus is on operating within the (legal) requirements for keeping all correspondence as part of project management; and
- ◆ A compliance/liability issue—focus is on finding out and addressing whether there are special rules for keeping email or if your organization is subject to the risk of subpoenas, which can lead to large-scale discoveries and disclosures.

Regardless of how an organization perceives the goal of its email management needs and responsibilities—or, for that matter, the way in which it needs to manage all of its vital information—its ability to develop and implement a technical solution can be a real challenge, particularly when taking into account the various needs across business units.

Reviewing Compliance and Liability

If compliance and/or liability are your organization's primary drivers for implementing an email management solution, you are probably aware of Congress's recently enacted Federal Rules of Civil Procedure (FRCP), which outlines the required procedures for civil suits in federal courts. Of particular relevance for email management are rules

26 and 34, which provide the guidelines for information discovery and disclosure in a civil context. In essence, an organization has the responsibility to make sure it is thorough in identifying and disclosing all possible sources and types of information that could be relevant to federal proceedings.

“Your organization needs constant access to, and control and awareness of, all of its electronic documents and email.”

With FRCP, if your organization is involved in a lawsuit, an overview of the electronic information you plan to use in the case must be made available before the formal discovery request. Your organization will also be required to take into account all possibly relevant data across every type of media and format, even if this data is stored remotely or by third parties.

Bottom line, your organization will need to have constant access to, and control and awareness of, all of its electronic documents and email that could be relevant to a civil case. Moreover, the processes that support these activities need to be in place before any litigation begins because trying to get control of your information “after the fact” will not only be very difficult but will also increase expenditures and organizational risk.

Thinking logically about adherence to FRCP, you can conclude that the following organizational activities need to take place in order to be positioned to meet FRCP guidelines:

- ◆ **Honestly evaluate** your organization's data storage and retrieval needs (in terms of security, data integrity, long-term storage viability and so on);
- ◆ **Establish appropriate systems** to make the searching, finding, retrieving and distributing of relevant data as simple as possible;
- ◆ **Create an organized and detailed overview** of where electronic documents are stored;
- ◆ **Assess document lifecycle policies**, particularly in terms of how well they can account for all your electronic information, especially email files; and
- ◆ **Ensure broad understanding and “buy in” of document lifecycle policies** and implementation processes for possible litigation-relevant information.

In sum, if you want to prevent a very costly investigation involving all PST files and email repositories in your organization (including backup tapes), you will need to have a proper filing plan in place, a digital platform to store your information, the ability to train your users and enforcement and audit mechanisms in place to ensure that the system is properly used. If you do not have these capabilities, you may be forced to turn over every piece of electronic information in your organization. One can only guess how much money this will cost!

Acknowledging Specific Compliance Issues

Given this compliance/liability context, implementing an email management solution

means you must be aware that at a certain moment in time you will be asked to find specific emails, review them and disclose them: the e-discovery and e-disclosure process. If you have ever been involved in one of these situations, you know that if you are not properly prepared, the cost of finding relevant messages, de-duplication, review, redaction and disclosure can be astronomical. By being properly prepared, you can save tremendous amounts of money and lessen your risk.

Below is a checklist for the types of performance capabilities that must be available with any compliance/liability-avoidance solution you integrate to manage your email. As you'll see in the next section of this article, many solutions on the market cannot offer all of these capabilities because the systems are based on proprietary, database-focused storage systems. Only open, non-proprietary XML-based systems, such as those offered by ZyLAB, can truly address all of the needs required by users needing flexible solutions to drive their email management needs.

The capabilities required for good email management that are fully supported by an XML framework include:

- ◆ **Capturing:** Make sure flexibility is built into your email capturing solutions, especially in terms of allowing individual users to archive emails directly from MS-Outlook in the proper sections of a filing plan. Also, the full capture of all email must be allowed from your email servers for auditing and enforcement purposes.
- ◆ **Storing:** Email needs to be stored in an open (non-database) format that is enduring and sustainable. You do not want to continually convert email collections over time.
- ◆ **Analyzing and enhancing content:** Make sure that automatic coding, text-mining, OCR (of bitmap attachments) and automatic translation tools are available to help you get 100% recall.
- ◆ **Searching:** Demand an advanced full-text search engine that can search terabytes of data.
- ◆ **De-duplication:** Find exact and near duplicates, which will often reduce the size of your data collection by 50% and save you time in the remaining phases.
- ◆ **Review:** Make sure that you can review documents quickly on their merit for the case and determine where redactions need to be applied.
- ◆ **Redaction:** Ensure that you can redact confidential, sensitive and private information.
- ◆ **Disclosure:** Rely on XML-based Web technology to securely disclose documents to other parties or burn relevant data to CD or DVD, which can lead to significant cost savings (such as on paper in extensive disclosures).

Thinking Outside the "Database Box"

As opposed to XML-based systems, proprietary storage systems embed a variety of risks. Consider, for example, the thousands of businesses in the 1990s that bought the popular optical storage systems. Not only were most of these systems proprietary, but the majority of the vendors in that segment have gone out of business or put their development and support resources elsewhere.

In general, the database-driven approach seems logical given many vendors' traditional or core expertise with database-oriented solutions. But being so attached to a traditional database-driven viewpoint automatically triggers hidden future costs when long-term email management is a requirement. For

"Being attached to a traditional database-driven viewpoint automatically triggers hidden future costs."

example, Microsoft SQL server has had four different versions over the last eight years, which means in most cases users must perform a conversion with each new release, as some of the older versions are not supported with newer versions of Microsoft's database.

The need to move forward and support all the considerations required by long-term, appropriate email management—security, sustainability, affordability (in terms of both upfront costs and minimal long-term expenditures for conversions or upgrades), compatibility (between systems, tools and storage media), and suitability for e-discovery activities—would seem to indicate building solutions less reliant on storage-focused database systems and more on systems built on, or at least in close association with, a flexible, enduring and affordable XML-based infrastructure.

In fact, some traditionally database-reliant vendors are acknowledging the need to move beyond reliance on databases, integrating some XML storage capabilities into their solutions. But making a token nod to

XML while still relying on databases to perform "heavy" storage doesn't really solve many of the core issues discussed here. Even a partial reliance on databases will still increase costs (as well as require upgrades every few years), create more opportunities for "liability vaults," enhance the potential that old information may eventually lose its integrity, pose potential integration issues and make detailed management of data, especially in proactive investigative situations, more difficult and risk-intensive.

XML-based, Compliance-modeled Email Management

Many agencies in the US government are using the types of processes advocated here, and are supporting them with XML, to great effect. These agencies are saving money and getting better performance and efficiency because their processes are thorough and their XML platform ensures them of the following:

- ◆ Sustainable, secure and enduring infrastructure, regardless of how much information is stored;
- ◆ Cost containment, due to no required conversions or high upfront costs; and
- ◆ An array of supporting tools that greatly enhance capabilities for e-discovery and e-disclosure.

Organizations can only hope to become more efficient and better positioned to avoid risks if they are able to honestly evaluate and develop the processes that they can actually implement and adhere to. As mentioned above, only a true, 100% XML solution can ensure all of the capabilities needed to support the email and document management processes now having to be implemented by so many organizations: cheap storage, enduring and sustainable, flexible and modular. At the same time, these solutions address functionality required in a legal discovery and disclosure process: searching, analyzing data, organizing, de-duplication, document review, redaction and advanced disclosure by Web technology, but also by CD and DVD. ■

ZyLAB is an innovative developer of affordable content management and compliance solutions for paper-intensive organizations. ZyIMAGE, ZyLAB's flagship solution, helps small and medium-sized businesses (SMBs) and government organizations digitally file and manage millions of pages of paper, electronic documents and email. High-quality search and retrieval features (which support over 200 languages) give users the ability to easily organize, investigate and distribute information.

With more than 7,000 installations worldwide and more than 300,000 users, ZyLAB has a wide breadth of experience and knowledge across a variety of different industries and business applications. For more information visit: www.zylab.com