



New Tools and Techniques Are Helping Organizations Lower the Risks and Costs of Managing Unstructured Data

*Richard F. Martin
Global Practice Director,
IT Service Management EMC Global Services*

Contents

New Tools and Techniques Are Helping Organizations Lower the Risks and Costs of Managing Unstructured Data	3
Why Is IM So Hard?	4
Extracting the Full Value from Unstructured Documents	5
New Software Tools Are Emerging	5
Specific Tactics Foster Success	6
Benefits: What Is the Result?	6
Execute IM in Phases and Include the Right Mix of Skills and Experiences	7

New Tools and Techniques Are Helping Organizations Lower the Risks and Costs of Managing Unstructured Data

If your organization isn't managing its unstructured information effectively, you're not alone. More than half of those surveyed in a recent EMC® research study admitted they were not classifying their unstructured documents—a critical first step to effective information management.

While companies may ignore the task, the risks and costs of doing so are growing every day. Unstructured file types—defined as anything not stored in a database—such as spreadsheet analyses, word files, pdfs, presentations, and audio and video files are becoming increasingly important in day-to-day operations. As much as 80 percent of corporate information is now in the form of unstructured documents. Many organizations are finding that key processes simply cannot be done without this information so its management has become more important than ever.

Federal and industry regulatory groups have taken note of that growing importance and the resulting increase in risk. New regulations like the Payment Card Industry-Data Security Standard (PCI-DSS) and the International Trade in Arms Regulations (ITAR) for defense contractors dictate how sensitive information like account numbers or design information is to be stored, accessed, distributed and secured. Legal discovery requirements mandate that an organization be able to produce information on demand or face substantial fines and sanctions. Over the past two years, it is estimated that over \$3B in fines have been levied against corporations that were unable to produce the right information on demand.

For other organizations, information management is a simple matter of economics. With the volume of these documents continuing to grow rapidly, the cost to store them is skyrocketing. Better management will help companies reduce required capacity by eliminating unused, duplicate or non-business related documents, improve accuracy by reconciling document inconsistencies, and improve efficiencies by ensuring that users are able to easily locate documents when needed and that documents are stored on the most cost-effective platforms.

As unstructured information is increasingly integrated with legacy applications within service-oriented architectures (SOA), it is imperative that companies begin addressing ways to better manage all information assets in an integrated, holistic way. The good news is that proper information management, starting with classification, forces the IT organization to become realigned with the new information needs of the business. Fortunately, a new class of tools for information classification and management is making it more likely for companies to tackle this challenge successfully.

Why Is IM So Hard?

Companies have spent significant time and money building tiered storage infrastructures to hold structured database information. With the volume of unstructured documents exploding, organizations must utilize these tiered storage infrastructures for unstructured information as well.

Management tools such as data movement and automated policy execution software that have been used successfully on structured data cannot operate on unstructured documents without effective classification of this information. What companies lack is an efficient means to classify these documents. Organizations that have struggled with unstructured document classification typically used a mostly manual process that produced very simplistic, static classifications, which were quickly outdated as the documents aged.

So how can organizations use these experiences to ensure success going forward? To start, it is important for everyone involved in the effort to have a clear and consistent understanding of the definition and objectives of information management.

Information management is a holistic approach to managing both structured and unstructured data that brings together previous independent efforts to manage structured and unstructured data such as RDBMS, ECM, Enterprise Search and Enterprise Portals.

Ironically, because they were implemented separately, these existing technologies have actually increased the barriers between structured and unstructured information and made it more difficult to accomplish the integration of all data types within business processes. Furthermore, none of these technologies addressed the challenges presented by manual classification, or provided an effective way to reflect how business process requirements affected the relationships between the unstructured documents and other elements of the information infrastructure.

Extracting the Full Value from Unstructured Documents

The fundamental problem that IM attempts to address is how best to leverage the value of the organization's combined information assets. Integration of the structured and unstructured data requires that there is an up to date inventory of all data. Since companies usually have a reasonable inventory of structured data, the first step involves identifying the unstructured documents that exist. IT, along with the lines of business, must determine where these documents are stored, who owns them, who uses them, which business processes require them, and the scope of their content. Then, the team must assess any related information policies that may already be in place by asking if these policies support the requirements and what policies should be in place but do not currently exist. By identifying any gaps between requirements and current information, the team creates a new set of business-based information policies that is used to classify all existing unstructured information assets.

New Software Tools Are Emerging

An emerging category of tools, called Information Classification and Management (ICM), promises to help companies get past the challenges of manual classification and has generated significant interest in this area.

Using these tools, which catalog both the attributes and actual content of files as well as their service level agreements (SLA) requirements, companies are able to keep documents classified properly as new data is created and existing documents change and age. Careful application of automation to the discovery and classification process helps ensure the ongoing change that occurs—documents created, versions updated, copied, deleted etc.—is accounted for, and that the relationships between them and the infrastructure are kept current.

Because these tools can assess both file attributes and actual content, organizations are able to orchestrate the actions, such as moving documents to a secure storage platform, that keep unstructured documents in compliance with corporate information policies with minimal manual effort and greater accuracy.

A key benefit of these tools is that they facilitate the federated integration of information that a service-oriented architecture requires. Instead of forcing a move of the documents to a consolidated storage platform, the tools provide all of the information needed to fully identify each document where it currently exists. Applications running within a service-oriented architecture (SOA) can use this information to access, process and distribute the documents as needed, and in compliance with corporate policies.

Specific Tactics Foster Success

Beyond these technology and methodology considerations, there are organizational and political challenges that companies must address to improve the chances of success of the IM effort. First, the lines of business should be closely involved in the classification process since the business process requirements must define the relationships between the unstructured documents, structured records, and applications. This helps ensure that any information policies that come out of the classification process are directly linked to the business model and support business process requirements.

Second, given the cross-functional processes that are served by IM, there are bound to be conflicts between functional organizations that only a senior executive would have the political capital to resolve. So, a senior executive should sponsor the initiative.

Having access to outside resources with experience utilizing the methodologies needed to map existing information assets and policies to business-based requirements and identify any gaps is also key. This experience gives the project team a politically neutral perspective that helps companies navigate the IM planning and implementation process effectively.

Benefits: What Is the Result?

Done properly, information management provides organizations with significant benefits. These include significantly lowered costs from better asset utilization; lowered risks resulting from better security and availability of critical unstructured documents; and compliance with regulatory requirements such as SOX, HIPAA, SEC, etc.

One financial services company was able to analyze 200 TB of unstructured documents and discover 10,000 unprotected files containing Social Security numbers. Using IM, the company classified these files as “sensitive” and applied the applicable storage policies, including moving the files to secure storage and encrypting the contents.

Companies will see much more efficient utilization of storage resources because IM makes it easier to classify data, automate policy-based actions, and meet service-level agreements while balancing infrastructure-related costs and service delivery based on the documents’ value to the lines of business. IM also greatly reduces the storage volumes required using techniques such as data de-duplication to reduce redundant documents and eliminate outdated versions where appropriate.

A *Fortune* 100 company used IM to reduce the costs associated with properly classifying the documents owned by employees who leave the company. By automating the process, the company was able to move many of these documents to lower-cost storage and save over \$6,000 for every exiting employee.

The emerging service-oriented architectures enable the bridging of applications and business logic across system and organizational boundaries. Legacy applications are being redeployed within SOAs so that users of a specific application can transparently access and use documents that are stored outside the scope of that application. In these new environments, all of the requirements that have been important to structured data, including managing both risk and costs, now apply to unstructured documents.

Because IM complements applications deployed in a service-oriented architecture, line of business users should see substantial productivity improvements as unstructured documents are delivered to users as needed in the context of the applications they already use. As a result, users make more timely, accurate and effective decisions.

Execute IM in Phases and Include the Right Mix of Skills and Experiences

IM is a comprehensive undertaking with wide-ranging implications, so organizations should avoid trying to do too much to start. A successful approach involves phases that start with the highest priority information first. The project team or sponsoring executive should make a decision up front to focus on either cost improvement or risk reduction objectives. This initial phase can also serve as a proof of concept that paves the way for subsequent phases of the IM effort.

The combined skills and experiences of the project team is another important factor in success. The project team should include experience with the technologies and methodologies that help optimize an information infrastructure. Project resources should also have practical experience helping other organizations utilize the latest automated software technologies such as discovery and classification tools, policy management software, and virtualization technologies.

Outside experience with IM will help organizations avoid the mistakes and challenges that held back previous efforts. Experience with compliance, continuity and security issues are important as these continue to grow in significance. Having the perspective of a proven, business-first approach based on industry best practices such as ITIL will ensure that the resulting IM tools meet the information needs of the business.



EMC Corporation
Hopkinton
Massachusetts
01748-9103
1-508-435-1000
In North America 1-866-464-7381

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC², EMC, and where information lives are registered trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.

© Copyright 2007 EMC Corporation. All rights reserved. Published in the USA. 05/07

EMC Perspective
H2707.1