

Content Management

Guidance on how to gain the most value from your greatest asset — your data

TABLE OF CONTENTS

- 2** Content Management = Lower TCO
- 2** Meeting Today's Data Needs
- 3** Document Management
- 3** Planning a DMS
- 4** Managing an Array of Content
- 4** Content Management Software
- 6** The Value of Content Management
- 6** Content Storage and ILM
- 6** SAN and NAS Solutions
- 7** Planning Your Content Storage
- 8** Manufacturer Options

Executive Summary

For many organizations, the data they possess is their greatest asset. For government agencies, this is especially true. Being able to reliably locate and access the agency's data is a necessity.

With a content management solution in place to efficiently capture, process, manage, store and deliver that data, organizations can fully leverage their content, utilizing it to further their operations and missions.

Unfortunately, not all government entities are getting the most out of their data. This is often because of an unstructured approach to managing their content. Isolated management systems put into use at different times by different departments have led to complex web of content islands.

Every operations application has its own process for managing information, making data sharing between applications difficult. Such an ad hoc approach to content management does not allow for the kind of organization-wide access to data that is needed by government agencies today. This lack of coordination costs organizations money, exposes them to risks and ultimately constricts the creation of new value from that information.

For organizations that lack a clear, concise approach to content management, the time for planning is now — enterprise content is estimated to be growing at a rate of up to 200 percent per year. This white paper will provide you with an understanding of the basic strategies and technologies that make up a content management solution, from capture to archiving.

Content Management = Lower TCO

IT departments are often asked to do the impossible. For example, organizations want a lower total cost of ownership (TCO) for managing their information even though organizational content is growing at an exponential rate (along with support staff costs to handle it). Or, individual departments want their unique content management needs met, while the organization wants better organization-wide sharing of content.

On top of these difficult demands, IT teams have to contend with many of the following content management issues:

- Higher TCO with each additional application added to the content stream
- Increased security and compliance risks
- Few comprehensive response options to legal requests and audits
- Organizational processes that are disconnected
- Poor information leverage because of limited access to and a lack of reuse of information

Many of these issues can be addressed by standardizing on a content management platform and aligning it with your organization's information infrastructure. A content management solution can easily and cost-effectively provide a broad range of solutions for your organization. It can also lower IT administration costs and improve the strategic use of information, while minimizing security and compliance risks.

A content management solution can provide common services such as security, workflow, version control, content retention and search no matter what form of content is being managed. An organization-wide content solution also helps decrease your dependence on specialized personnel to maintain the application environment.

With a single depository for content, departments can easily share and reuse their content with other areas in the organization. You'll also be able to safely share your content with outside organizations including partners, vendors and the public. All of these pluses make for a more efficient organization and lower the TCO of managing your content.

Meeting Today's Data Needs

Today's operations environment produces more documents, information and knowledge than most government agencies can handle. Storing, managing and retrieving all of this data has emerged

Five Components of Content Management

A content management strategy can be broken down into five basic actions:

- **CAPTURE:** Organizations collect paper documents and electronic documents using image capture, optical character recognition, forms processing and data collection tools. A scanner may be required to transform paper documents into pixels.
- **MANAGE:** This component involves the use of databases and specific applications and solutions such as document management, collaboration, web content management, records management and workflow. An organization can link these tools so that it's possible to access data across applications.
- **STORE:** Data repositories, library services and storage technologies are an essential part of content management. These include magnetic tape, optical media such as CDs and DVDs, and disk-based storage devices such as redundant array of independent disks (RAID), storage area networks (SAN) and network attached storage (NAS).
- **DELIVER:** When users require information, it's essential to put it at their fingertips. Enterprise desktop search tools such as X1 and Google Desktop Search are a start, but it's also necessary to provide file viewers, file converters, security tools (including encryption) and distribution mechanisms such as intranets, portals and mobile devices to make information delivery quick and seamless.
- **PRESERVE:** Backing up and archiving information are essential parts of content management. An organization can turn to a variety of tools, including optical disks, tapes, storage devices based on SAN and NAS solutions, and microfilm.

as an enormous task, and one that requires a serious commitment to staff, equipment and funding.

Without systems in place to handle all the content and deliver the right data to the right place, it's next to impossible to achieve the level of productivity and fiscal responsibility required of our government agencies.

Content management attempts to remedy the problem by handling a diverse array of tasks, including the capture, storage, retrieval,

distribution, preservation and destruction of documents and content. These include everything from written reports in Microsoft Word and figures residing in an Excel spreadsheet to e-mail and instant messaging strings.

Content management can also extract data from large databases and cull through unstructured multimedia files, including video, audio and photos. Although managing content has always been a key consideration for government agencies, the enormous growth of data, along with more complex and sophisticated storage and retrieval needs, has drastically altered the equation over the last few years.

Regulatory requirements, legal challenges and burgeoning knowledge management initiatives have all fueled a need for retaining and finding records. In addition, a growing number of organizations must provide information to customers and constituents — often through a website or portal. The ability to cull data and present it in an appropriate manner is paramount.

Yet, achieving success with content management is not so simple. It covers all kinds of tasks, from web content management to fixed assets such as scanned documents. There are content management applications to examine, hardware and software considerations to ponder and workflow to analyze.

The most successful organizations also understand that content management is constantly evolving, and that it presents an array of unique challenges and opportunities.

Document Management

The first step to managing content is getting the organization's information into a digital format. This can best be handled with a document management system (DMS).

A DMS is a collection of integrated tools that converts paper documents into electronic form to be stored, shared and easily accessed. It can help streamline everyday operations processes and improve operational efficiencies.

Document management products fall into three categories: document imaging systems, electronic document management systems and integrated document management systems. Document imaging tools, such as scanners and multifunction devices, and the specialized software that lets users capture, manipulate, index and store images, serve as the front end of document management systems.

Among applications, Adobe Acrobat — familiar to anyone who has ever read a PDF file — has long been a market leader. This document

management tool enables staff, using Adobe Acrobat Connect software, to collaborate in real time over the web.

Adobe Reader lets users electronically fill out, digitally sign and locally save an electronic form and its data. The Acrobat software family consists of Adobe Acrobat 8, Adobe Acrobat 3D, Adobe Acrobat Capture 3.0, Adobe Acrobat Connect, Adobe Document Center and Adobe Reader.

Other top tools include FileBound, which emphasizes records management; Kofax Capio, which captures and stores documents as PDF files; and McAfee Email Compliance Module, an optional component of the McAfee Secure Messaging Gateway and Secure Internet Gateway appliances, specializing in helping organizations meet regulatory guidelines.

On the hardware side, scanners range in speed from Xerox's DocuMate 510, a low-cost unit that scans 10 pages per minute, to Fujitsu's fi-5900C, an industrial-strength device that can scan up to 100 pages per minute. Some scanners, such as Epson's Perfection 4990 Photo Pro and Canon's CanoScan 4200F Color Image Scanner, are more suited for scanning photographs and other images. Multifunction devices, such as the HP LaserJet 4345x, let users print, copy, scan to e-mail and fax.

Planning a DMS

Before implementing a new DMS, it's beneficial to understand where your organization is right now. In addition to evaluating current processes and your overall IT environment, survey staff members to help identify your requirements. Be sure to define the amount of information you need to manage and map out your current workflows, as well as how your files will be indexed, routed or archived.

A well-defined requirements document that includes workflows will help you choose the right technology for your document management system. The more time you spend developing your requirements, the better your chances are of creating a system that meets your needs. Make sure that everyone who will be impacted by the project is on board before implementation.

With requirements and architecture documents in hand, you can begin to compare the available technology solutions and choose the products that best fit your system. Once you've determined your needs and chosen your products, you can begin assembling your DMS. Consider starting with a pilot program. You can then tweak your system as needed, based on the results of the pilot program.

Managing an Array of Content

As more data is generated and retained and databases grow larger, organizations need to access paper documents and other unstructured data that typically reside in file folders, hard drives and other storage media. Finding the right data when you want it has become one of the vexing challenges of the digital age.

For organizations that rely on manual keystroke entry to add paper documents to electronic databases, the process is rife with errors, delays and high costs. Make no mistake, the transition to a paperless office isn't easy. Content management tackles the problem by using an array of tools and technologies to consolidate and streamline these processes.

For example, an organization might use a high-speed scanner to enter paper-based forms into an application or database. As a result, it is able to eliminate manual entry of the data.

But once the images or data fields exist within an application, it also needs to store the data, index it and allow users to retrieve it on demand. It also might need to mingle scanned data and web pages with electronically generated forms and fields in a database to provide a complete picture to administrators or analysts.

Because content management can encompass so many areas, it's crucial to pay attention to several core areas, including document management, collaboration, rich media management and records management.

In fact, unstructured data is taking on greater importance all the time — and it now accounts for approximately 80 percent of all content. There's been an explosion of data that doesn't fit into the neat rows and tables of a database — data that a computer cannot easily act on and process.

As a result, it's essential to understand where the data resides and how the organization can tap into it — usually through powerful search tools built into a content management application such as EMC's Documentum or Oracle's Stellent Universal Content Management, or a powerful stand-alone search tool such as X1.

The latter indexes a hard drive or a network and provides near instant results across a wide array of parameters, including e-mail, attachments, files, pictures, music and contact management applications such as Outlook.

There's no single approach to content management. Each organization must tailor an initiative to support its unique needs and requirements, whether it's to meet organizational objectives or ratchet up compliance to regulatory initiatives.

The Spin on Solid-state Drives

As organizations collect more content, accessibility becomes an issue. The more data there is, the more there is to comb through to find what you're looking for, the longer the process takes. The latest flash solid-state drives (SSDs) utilize flash memory, which operates much faster than any spinning hard drive on the market.

SSDs offer the potential to consolidate vast volumes of network storage while providing continued real-time access. These drives can be extremely cost effective when used for databases, data warehouses and scientific applications. This is because of their higher memory bandwidth (typically 100 times or more than a typical drive).

SSDs can benefit organizations that have a lot of servers that are underutilized. They require less cooling and use less energy than standard hard drives. Because SSDs can reduce the number of servers in use, the number of software licenses an organization needs goes down. While SSDs are pricey (though prices are dropping), the cost savings they offer can make them very attractive to organizations.

Moreover, it's essential to choose the right applications and equipment to streamline all the various functions and tasks. For example, a federal agency might receive both paper-based forms and electronic forms and need to store both together in a database. This might require automated scanning on the front end and a powerful search feature on the back end.

Without the ability to pull out the data — and associated e-mail messages, faxes and database entries — staff time likely would skyrocket and productivity would plummet. Similarly, an agency might need to cull data residing in different storage devices, servers and systems, and present it through a website or government e-portal.

Content Management Software

Content management isn't a particularly new idea. However, as technology and processes have become more sophisticated over the last few years — and web-based technology has taken root — the ability to manage content has grown exponentially.

The appeal of content management isn't difficult to understand. A successful implementation boosts productivity, trims costs,

cuts cycle times and ratchets up collaboration and the exchange of knowledge. By arming staff with the ability to improve critical thinking and decision-making, an organization can greatly improve its performance.

But content management is more than the sum of its parts. With the right systems and processes in place, it creates a level of synergy that transforms an organization.

Today, the debate isn't whether an organization should use content management (it should), but how to approach the challenge of

Assessing the Content Infrastructure

For organizations looking to bring some unity to their content management approach, a good place to start is with a thorough assessment of the current content management infrastructure. This can be done by gathering information through the following steps:

1. Find out how many different content management applications are in use either by the IT team or individual departments.
2. Determine if there are plans in place to deploy additional content management applications and if the IT team will be supporting them.
3. Identify which departments need access to the application.
4. Figure out how many people need to support each content management application. Use this figure to estimate the personnel costs to support the existing content management infrastructure and integrate new applications.
5. Identify which departments have the greatest need to share information across the organization.
6. Find out what processes span multiple departments.
7. Determine how information is being shared across the entire organization, between departments and within departments.

Gathering all of this information about the current content management infrastructure will help you see where the inefficiencies are and provide some guidance on the direction you will want to take in implementing a content management solution across the organization.

managing data. Numerous manufacturers offer content management software solutions, including EMC Insignia, EMC Documentum, Filebound, Laserfiche, Questys Solutions, Workshare and X1.

Not surprisingly, each solution handles processes and workflows in a different way. EMC's Documentum, for instance, offers a high degree of flexibility and scalability. Using a single repository, it manages virtually all types of content across multiple departments. This "unified" approach makes it easier for groups to share and reuse content with other departments and divisions within an organization.

Documentum also facilitates secure sharing with outside organizations, including partners, vendors and the public. It accommodates a variety of content types, including HTML and XML, graphics, multimedia, other types of rich media and traditional documents created with desktop applications.

And it provides an embedded search engine that indexes and retrieves documents from within a single interface. This feature eliminates the problem of searching different systems and using different tools.

Some applications, such as Laserfiche, not only store data and provide tools for finding it, but also offer audit trails and compliance tools. Others, such as Workshare, provide sophisticated document security and controls to ensure that only staff with authorized rights to view and edit documents are allowed to do so.

Still others, including Questsys Solutions, provide powerful tools for managing documents, records and assets as well as web content. It can manage thousands of websites and allow an organization to improve the authoring and editing processes.

Search also has grown into a core consideration within the content management software umbrella. X1, which helps users sort through information with lightning-quick efficiency, is an example of how far the search function has come. It uses a single interface to access and feed information from across a wide array of applications and locations.

What's more, X1 offers robust tools that let IT administrators control data access so that it matches existing security policies within the organization. This approach cements the relationship between content management and information access.

It's no small matter. When a content management software solution is used effectively, it imposes additional structure and context around information. The benefits come from increased visibility and reuse of information, as well as clarity about which documents exist and where they came from. Beneficiaries range from staff

gaining improved access and visibility to information, to IT, legal and compliance organizations that gain single-source access to critical and sensitive data.

The Value of Content Management

Building a framework for operations success requires an ongoing commitment of time, attention and budget. An organization that is looking to take a best-practices approach to content management will be able to manage diverse data, information and knowledge, and integrate everything into its workflow.

A content management solution is far more than a technology equipped with a powerful search engine; it's an "organized repository" that incorporates library services (including check-in and check-out records), version control and document-level security.

When used effectively, it eliminates paperwork, reduces the need for storage, offers faster access to information, delivers streamlined processes, provides an audit trail and boosts overall security.

Going forward, the demand for a focused approach to content management will continue to grow, and new uses for the technology will emerge. Over the next few years, content management is likely to become an essential tool for delivering highly targeted content to

PDA's and mobile phones, and could play a core role in expanding online collaboration tools, websites and enterprise portals.

It is also likely to make the paperless office a reality. Going forward, content management will play a larger role at the center of today's digital workplace.

Content Storage and ILM

Content storage — an integral part of any content management solution — is rapidly growing as a percentage of the total IT budget, with some organizations allocating 40 percent of their expenses to storage hardware, software, administrative costs and maintenance. This upward tick has led to many organizations rethinking how they handle their content storage needs.

The old way of dealing with expanding content storage needs was to just add more capacity, usually in the form of high-performing (and high-cost) hard drives and storage arrays. And why not? For years, the cost-per-megabyte of all types of storage hardware and media has been steadily dropping.

But today's budget-conscious IT departments need to find a more cost-effective strategy to adequately address their expanding content storage needs. A more sophisticated strategy is required to keep storage costs under control, to make sure organizations have the storage capacities they require and to provide for optimal access, backup and restore services.

Many organizations are opting for a holistic approach to storing their information called information lifecycle management (ILM). This approach to managing content storage is based on the idea of categorizing information according to its value and age.

Tiered storage (a part of an overall ILM strategy) helps organizations allocate their storage resources more effectively. Slower, but still highly reliable drives, including parallel ATA (PATA) and serial ATA (SATA), hold information that's key to an organization's operations. But because this information is called upon only sporadically, it can exist on slower-performing equipment.

Information held in archives for legal and regulatory compliance, which is rarely accessed, is best housed on low-cost tape cartridges or, in some cases, ATA drives.

SAN and NAS Solutions

The next step is to make sure you're using your content storage assets at their optimum levels. Agencies routinely use only 25 to 40 percent of their total storage capacity because the resources are segmented within individual departments and workgroups.

Storage Virtualization

Virtualization technology is making its way to content storage as well. Storage virtualization extends the NAS utilization concept to span an entire organization. Government agencies can manage their storage investment from a central vantage point rather than as individual NAS or SAN systems.

Some storage manufacturers offer virtualization applications designed with the assumption that end users operate heterogeneous storage environments, with storage hardware from multiple vendors based on price and performance considerations.

Virtualization software that accommodates a wide range of products offers flexibility in pooling storage and migrating from one device to another. With the right virtualization plan in place, real-world utilization rates can easily rise from less than 30 percent to greater than 70 percent or even more.

Underutilized storage space is expensive not only because organizations aren't taking advantage of everything they're paying for, but also because of indirect costs for resources such as physical space and electricity that are used by these storage resources. Inefficient content storage management can have a cascading effect in terms of wasted resources, so it's important to manage your storage as efficiently as possible from the start.

Two alternatives to combat storage underutilization are network attached storage (NAS) and storage area networks (SANs). Rather than using direct attached storage (DAS) or individual drives connected to every server, users have the option of connecting storage directly to a Fibre Channel fabric — a SAN — to create pools of storage capacity that can be shared by a number of servers.

This allows users who have files strewn across a litany of servers to consolidate their file and print services onto a "storage server" or NAS appliance. In both cases, benefits include the ability to centrally manage content storage and to add drives to the network as needed without having to shut down any associated production servers, a common occurrence with DAS.

Planning Your Content Storage

Content storage-savvy organizations can follow a three-step approach to better manage their content storage needs.

STEP 1: DEVELOP A FORMAL STRATEGY THAT ADDRESSES YOUR CONTENT STORAGE ISSUES. The key to management is to know what you have in terms of data and information. Many organizations only know they have a certain amount of gigabytes or terabytes of storage.

A best-practices approach requires organizations to dig deeper: How much of that content is duplicate information, how much is nonessential data, and conversely, how much is absolutely critical to your operations? You can't properly manage what you don't have a full understanding of.

Storage resource management tools look at all the data throughout the network and compile reports about usage, file types, duplication and other characteristics. These tools can also give organizations a clear picture of the type and amount of the various storage technologies that are currently being devoted to direct attached and network attached storage systems, as well as to storage area networks.

EMC's Visual SRM (Storage Resource Management) solution allows organizations to take a snapshot of the existing content storage

Stimulating Storage

Federal and local government healthcare agencies will soon be grappling with many of the major issues related to content management and storage. The recently passed American Recovery and Reinvestment Act includes funding for healthcare organizations to implement electronic health records (EHRs).

"The move to EHRs will exponentially increase the amount of data that is digitized," says Brian Babineau, a senior analyst at storage consultancy Enterprise Strategy Group. "That will cause storage demands to go way up."

As a result, healthcare organizations should start thinking about implementing a central storage environment.

"An EHR is essentially a combination of many different files; and to be able to make sure all of those individual files can be managed as an entity, the storage has to be shared," Babineau says. The best way to do that is to set up centralized storage where all applications are saved in a single location.

environment. It records storage resources, files and database systems; it gathers information regarding size and utilization levels of files; and it can provide the usage attributes of that information. With this data, organizations can set policies to move data to the right place at the right time.

Other types of automation software, such as EMC Legato's EmailXtender, address more specific problems. EmailXtender allows users to set policies to mailboxes within a messaging system and determine when information is moved from high-performance storage access to slightly lower-performing, more economical storage assets. Veritas' Enterprise Vault e-mail archiving software integrates with Microsoft Exchange and provides policy-based archiving tools.

STEP 2: MAP OUT A STORAGE STRATEGY. From the management tool reports, IT administrators can begin categorizing the data to develop their ILM plans. Particularly important is the ability to determine the cost-per-megabyte the organization spends for its storage resources, a number that becomes the benchmark for gauging the success of subsequent cost efficiency efforts.

Administrators can rank data by considering a hypothetical system crash. If there is a disaster, what information would the organization

need up and running within milliseconds or seconds, and what information is essential but not necessarily something that will be needed immediately after a failure?

IT departments must also work with organizational leaders and legal departments to help determine where to place the data. They must decide how critical the data is, how long the data must be kept and how quickly they need to access it. For example, retrieving information on disk is faster than using tape.

Making the process more complex is the fact that data can have different values to an organization at different times. If an organization is served with a subpoena that requires it to unearth five years of e-mails on a certain subject, that data suddenly becomes important. While some data will always be valuable, some whose value has decreased can increase again.

By becoming more efficient in terms of how it uses content storage resources, an organization may be able to delay hardware purchases or consolidate existing resources to save money. Also, when administrators have a sense of all of the different levels of data, they can buy less-expensive solutions for lower-tier items, rather than bearing the expense of treating all data like top-tier information.

Storage analyses like these provide another cost-saving opportunity — reducing resources needed for backups. By not backing up duplicate or unnecessary info, back up is quicker and easier.

STEP 3: TARGET INVESTMENTS IN HIGH-PERFORMANCE DISK ARRAYS, LOWER-PERFORMANCE DISKS AND TAPE DRIVES. By successfully matching the right storage alternative to each class of data, organizations don't waste money for performance overkill on less-critical data, nor do they frustrate users by not giving them access to important information when they need it.

For most tiered storage operations, tape remains an integral part of a tiered strategy. Thanks to newer and more efficient compression technologies, the cost-per-megabyte for tape systems can drop to a penny or less.

However, inexpensive ATA drives are making some inroads into the once tape-dominated backup market. Drive-to-drive backups, while more expensive, offer the advantages of RAID (redundant array of independent disks) redundancy to provide an extra measure of assurance against data loss.

Disk vendors are also building ATA-based RAID arrays that match one controller for each disk, which increases reliability and achieves faster data-transfer rates. This allows an organization to have zero shared traffic, which avoids "collisions" as multiple drives vie for controller bandwidth.

Making this one-for-one disk/controller option easier for hardware suppliers are the increasingly popular SATA drives, which are replacing the traditional parallel ATA interface devices. SATA drives also depart from traditional ATA drives by offering more of the performance monitoring and support functions common in costlier SCSI devices.

Manufacturer Options

EMC has solutions (both software-based and hardware-based) geared toward many aspects of content management. The Documentum family of products can help organizations build and maintain a content management system, from the basic platform to development tools for customized content applications.

HP offers LeftHand and StorageWorks storage area network solutions for organizations. The LeftHand SAN solution works well in virtualized server environments and is optimized for database and e-mail applications. The StorageWorks solutions cater to shared storage situations that utilize either a file server or an iSCSI target for application/virtual environment storage, or both.

LASERFICHE manufactures an enterprise content management suite that covers document management, records management and operations process management functionality. Some of the tools available with this solution allow organizations to scan, index, structure workflow, and integrate, distribute, track and audit content.

QUESTSYS SOLUTIONS specializes in document and content management software. The content management suite helps organizations with data capture, storage, indexing, search and retrieval, access management and output and distribution.

X1 TECHNOLOGIES specializes in indexing and search solutions for digital content. This manufacturer has solutions geared specifically toward e-mail search and application-specific searches for applications such as Microsoft SharePoint and Oracle.