Document Management Overview

Contents

I. Introduction: Streamling Business Processes through Document Management 1

Benefits of Document and Records Management.......... 2

II. Document and Records Management Defined.......... 3

Document Management.......... 3

1. Capture, or the Ability to Import Different Types of Documents.......... 3
2. Storage and Archiving That Allow for Growth and Change.......... 5
3. Indexing and Retrieval, or the Ability to Find What You Want When You Want It.......... 7
4. Distribution, or Putting Information in the Hands of the Right People.......... 8
5. Security, or the Ability to Protect Your Documents from Loss and Tampering.......... 8

Records Management, or Documenting Business Activity.......... 9

III. Essential Components of Document Management Solutions.......... 10

Usability.......... 10

Capture.......... 10

Annotations.......... 11

Storage and Archiving.......... 12

Distribution.......... 14

Security.......... 15

Integration.......... 17

Technical Considerations.......... 17

Records Management Applications.......... 19

IV. Implementation: Addressing Your Business Needs.......... 21

Records Management Considerations.......... 22

Scaling from Pilot Project to an Organization-wide Solution.......... 22

Training.......... 22

Support and Maintenance.......... 23

Outsourcing Scanning.......... 24

Compliance and Legal Issues.......... 24

Contents

V. Frequently asked Questions 26
General 26
Capture 27
Indexing 28
Viewing/Printing/Exporting 30
Records Management 31
COLD (Computer Output to Laser Disc) 32

VI. Glossary of Terms 33

I. Introduction:

Streamlining Business Processes through Document Management

I“ve done searches that would have taken me probably three or four working days, and I found the information in about 10 minutes. Our archives are historical treasures Ð which is one of the reasons we did this, because people use them for research and the records were wearing out. So we wanted to store the original materials away and not risk damaging them anymore.Ó

,

Streamlining business processes and increasing productivity are fundamental concerns for any organization Ð private, public and non-profit alike. In an increasingly strict regulatory environment, managing documents and records diverts significant time from an organization's mission-critical objectives.

Document and records management software has many benefits that can appreciably improve organizational efficiency. Since these applications are complex systems that represent a solid investment, organizations should carefully evaluate their current and future needs beforehand.

ColorNet created this guide to give organizations perspective on document and records management systems and on the demands of implementation. ColorNet provides this work as an educational resource. It results from our nearly twenty years of experience helping customers solve their business problems and represents our commitment to educating organizations and individual users about the technology of document management.

This guide is divided into parts that clarify different aspects of document management. The introduction outlines the broad business benefits of document and records management systems. Explanations of the basics of document and records management systems follow. Specific system components crucial for improving business processes are then detailed. The key elements for successful implementation of a document management system are discussed in the next chapter. Frequently asked questions and a glossary of terms related to document imaging, document management and records management are located at the end of the guide.

Benefits of Document and Records Management

Document management systems are software applications that capture paper and electronic documents and provide the storage, retrieval, security and archiving of those documents. Records management is a specialized discipline. In particular, it is a set of recognized practices related to the life cycle of records Ð information that serves as evidence of the business activities of an organization.

The document management process begins with the conversion of paper documents and records to electronic files. Digitizing eliminates the many obstacles created by paper Ð labor-intensive duplication procedures, slow distribution, misplaced originals and the inconvenience of retrieving files from remote locations. Because paper files are also costly to process, duplicate, distribute and store, digitizing reduces operating expenses and overhead.

Document management applications enable more efficient distribution of and control over information, files and records throughout the organization. These software programs simplify business processes by automating repetitive procedures, document routing and e-mail notification. Document management systems expedite business processes by allowing instant access to information; greater collaboration within and among departments and offices; enhanced security for files and records; and the application of procedures that facilitate compliance with record-keeping requirements imposed by the SEC, NASD, HIPAA, Sarbanes-Oxley and others.

Document management makes it possible to:

¥ Manage millions of documents and retrieve the right one in seconds.

¥ Share documents with colleagues while protecting confidential information.

¥ E-mail and fax files instantly.

¥ Access documents while traveling.

¥ Publish documents to CD, DVD or the Web, as appropriate.

¥ Back up files and records for disaster recovery.

Records management systems simplify the life-cycle management of business records. A records management system supports the automatic enforcement of consistent, organization-wide records policies and reduces the cost of regulatory compliance.

Records management software provides:

¥ Improved efficiency in the storage, retention and disposition of records and record series.

¥ Detailed reports of which records are eligible for transfer, accession or destruction.

¥ Audit trails to track all system activity and the entire life cycle of records.

II. Document and Records Management Defined

Document Management

Document management begins with the conversion of paper or other documents into digitized images. These images can be easily organized and quickly retrieved, indexed and archived. When files are scanned or electronically converted, a high-resolution digital copy is stored on a hard drive or optical disc. Templates, or electronic index cards, can attach information, such as author, reference number, date created, or key words to a document. Files can still be viewed, printed, shared and stored. Which documents people can read and what actions they can perform on these documents depend on the level of security that the system administrator has assigned to the user.

All document management systems should have five basic components:

¥ Capture for bringing documents into the system

¥ Methods for storing and archiving documents

¥ Indexing and retrieval tools to locate documents

¥ Distribution for exporting documents from the system

¥ Security to protect documents from unauthorized access

Digital document management represents a significant advance over storing information on paper. No longer just ink on a page, the document becomes active content after being processed by Optical Character Recognition (OCR) technology. A document management system should offer effective search tools for document retrieval, including full-text search, index field searches and a visual filing scheme that permits users to browse for documents.

Following is a description of the five basic components to look for when choosing your system.

1. Capture, or the Ability to Import Different Types of Documents

There are three primary methods of bringing files into a document management system:

¥ Scanning or imaging, for paper files

¥ Importing, for archiving electronic documents such as word processing files, spreadsheets, faxes, audio and video

¥ Conversion, for creating unalterable images of electronic documents

Scanning

Scanning a document produces a raster (picture) image that can be stored on a computer. When you choose a scanner, it is important to consider the size and volume of paper to be scanned, along with price and overall budget. The ability to support a wide range of scanners is one of the defining characteristics of a versatile document management system.

A scanner should have an Automatic Document Feeder (ADF). The ADF speeds up the scanning process by allowing stacks of paper to be placed into a tray and automatically fed one page at a time into the scanner. Scanners without an ADF require each page to be manually placed in the scanner; they are designed primarily for imaging graphics.

Scanners can handle a variety of paper sizes, from business cards to engineering drawings. Most departments only need to scan documents up to legal-size paper ($8^1/_2$-inch x 14-inch). For organizations or departments that use blueprints, building plans and architectural drawings, there are large-format scanners that support E-size (34-inch x 44-inch) documents. In general, the larger the paper size the scanner can handle, the more expensive it is. Other options, such as color or grayscale, also increase the scanner's price.

The speed of the scanner is another consideration. Document imaging scanners can handle between 10 and 200 pages per minute. These are available in both simplex mode and duplex mode. Duplex scanners allow both sides of a two-sided document to be scanned in a single pass. High-speed scanning and duplex scanning can increase the price of the scanner. In some instances, it is more economical to purchase two 20-page-per-minute scanners than one 40-page-per-minute scanner. However, the two-scanner option is only supported by document management systems that support multiple scan stations.

Importing

Document importing is the process of bringing electronic files, such as Microsoft® Office suite documents, graphics, audio clips or video files, into a document management system. Files can be dragged into a document management system and remain in their native formats. These files can be viewed in their original format by either launching the originating application or by using an embedded file viewer from within the document management system.

Conversion

Converting documents is the process of transforming electronic files, such as word processor or spreadsheet documents, into permanent, raster-image format for storage within a document management system. Windows applications, such as Microsoft Word, Excel or Autodesk AutoCAD, can print existing files into an unalterable image of the document. These images are usually stored as archival-quality TIFF (Tagged Image File Format). For documents, the conversion process also pulls a clean stream of text directly from the document, eliminating the need for OCR. This text file can then be used for full-text indexing of the document to assist with later retrieval. Converting electronic documents bypasses scanning, saves paper and printer ink and produces a cleaner image than scanned paper files. The document management system should be integrated with Microsoft Office or other applications to permit users to convert documents with maximum ease. This method of imaging electronic documents is best suited for permanent archives.

2. Storage and Archiving That Allow for Growth and Change

Once brought into the system, documents must be reliably stored. Document management storage systems must be able to accommodate changing technologies and an organization's future growth. Hardware independence is critical to assuring that a document management system will meet all of your current and future needs. A versatile document management system should be compatible with all storage devices currently available – as well as those on the horizon – to provide long-term document storage or archival.

To ensure the future readability of documents, a document management system should store files in a non-proprietary format, such as TIFF or ASCII. Storing document images or text files in a proprietary format may leave your organization dependent on the future success or failure of another company.

Currently, there are five primary storage options:

- Magnetic Media (Hard Drives)
- Magneto-Optical Storage
- Compact Discs
- DVDs
- WORM

The advantages and drawbacks of each are described below.

Magnetic Media (Hard Drives)

Increasingly fast response times to store and retrieve a document, along with dramatic reductions in storage prices, make magnetic media a popular choice. These systems include Redundant Array of Independent Disks (RAID), Network Attached Storage (NAS) and Storage Area Networks (SAN). These devices are relatively inexpensive, can be linked together to store large numbers of documents and provide fast response times.

The main drawback of magnetic media is that, while inexpensive, they still contain moving parts, which are subject to mechanical failure. Data files can also be completely erased. Computer personnel should perform regular backups of hard drives so that if data is erased or damaged, it can be restored.

Magneto-Optical Storage

In the past, the magneto-optical (MO) diskette/disk drive was a popular way to back up files on a personal computer. As the term implies, an MO device employs both magnetic and optical technologies to obtain ultra-high data density. A typical MO cartridge is slightly larger than a conventional 3.5-inch magnetic diskette and looks similar. But, while the older type of magnetic diskette can store 1.44 megabytes (MB) of data, an MO diskette can store many times that amount, ranging from 100 MB up to several gigabytes (GB).

The chief assets of MO drives include convenience, modest cost, reliability and, for some models, widespread availability approaching industry standardization. MO disks can be placed in jukeboxes that hold hundreds of disks. The chief limitation of MO drives is that they are slower than hard disk drives and still subject to mechanical failure. Data files can also be completely erased. With the drop in the price of hard drives, the popularity of magneto-optical storage has faded.

Compact Discs

Compact discs (CDs) are small discs used to store digital information. Since nothing touches the encoded portion of the disc, the CD is not worn out by the playing process. Standard CD formats include CD-ROM (Compact Disc-Read Only Memory), a preprinted media format; CD-R (CD Recordable), a single-use recordable disc; and CD-RW (CD Rewritable), a multi-use recordable disc.

CDs offer a safe and reliable medium that can provide long-term storage for images. Moreover, CD-ROMs do not require specialized hardware or software to retrieve information. CDs use ISO-9600 specifications; this means the data can be read on many computer platforms. The primary drawback of this medium is its limited storage capacity, 650 MB. CD-ROMs can be accessed through CD-ROM drives, CD towers and jukeboxes of up to 500 discs, making it a convenient method of storing large numbers of imaged documents.

DVDs

DVD, which stands for Digital Video Disc or Digital Versatile Disc, is another form of optical disc storage technology. It is essentially a faster CD that can hold more information, including video, audio and computer data. DVD aims to encompass home entertainment, computers and business information within a single digital format, eventually replacing audio CD, videotape, laser disc, CD-ROM and even video game cartridges. DVD has unprecedented, widespread support from all major electronics companies, all major computer hardware companies and about half of the major movie and music studios.

Since the disc is read by a beam of laser light, there is no wear and tear, even if it keeps rereading the same data. The tough plastic surface is forgiving of fingerprints, dust and dirt. This means DVDs can be played thousands of times and continue to represent the best long-term option for reliable document management storage. The drawbacks of this medium are its high costs and an ongoing standards battle at time of publication, as different manufacturers are using different formats for rewritable DVDs.

WORM

WORM, which stands for Write Once, Read Many, is an optical disc technology that allows you to write data onto a disk just once. The data is permanent and can be read any number of times. This format is not readily available and requires specialized hardware and software to operate. Unlike CD-ROM, there is no single standard for WORM disks, which means that they can only be read by the same type of drive that wrote them. This has hampered their acceptance, although they have found a niche market as an archival medium.

While this standard definition of WORM refers to a specific type of storage technology, WORM has taken on a broader meaning in other contexts, such as financial services, to include any optical disc that is, in practice, a write-once-read-many medium. In this general sense, WORM includes more common storage media such as CDs and DVDs.

3. Indexing and Retrieval, or the Ability to Find What You Want When You Want It

A full-featured document management system makes retrieval of relevant documents fast, easy and efficient, and offers multiple methods of indexing, or categorizing, information. Indexing allows users to quickly sort large volumes of data to find the right document. Whatever the combination of indexing methodologies, search methods need to be easily used and understood by the people who retrieve the documents, as well as those who file them.

There are three primary ways of indexing files in a document management system:

- Full-text indexing, or indexing every word contained within a document
- Index fields, or indexing through keyword categories of documents
- Folder/file structure, or indexing by associated document groups

Retrieval is where the quality of the indexing system is most evident. Some document management systems let users search only by indexed keywords, which requires a person to know how the document was categorized and what index fields were assigned to it. A powerful indexing system will make it possible for users to find any document based on what they know, even if that amounts to no more than a word or phrase within the document. The more a document management system adapts to an organization's existing procedures, the less upheaval and training are involved for users of the system and the greater the likelihood the system will be used on a regular basis.

Full-Text Indexing

Full-text indexing allows users to locate any word or phrase that appears in the document. By providing full-text indexing, document management systems can eliminate the need to read and manually index documents using keywords.

To enable full-text indexing, the software must have the capability to perform Optical Character Recognition (OCR). The OCR process translates printed words into alphanumeric characters with near-perfect accuracy, enabling each occurrence of a word to be tracked by the application. OCR dramatically reduces the cost of manual indexing while providing improved search capabilities.

However, OCR cannot process handwriting or images. Moreover, when a computer performs OCR on a document, it typically uses English as the default alphabet. If multiple languages are required, the document management system should support OCR and full-text searches in these languages. To avoid creating extra work, a well-designed document management system should provide the ability to automate the OCR and full-text index processing of documents.

Index Fields

Index field searches enable users to comb through millions of records in seconds to find necessary documents. The ability to use index field information to locate documents is important in cases where a topic search is more expedient than finding every occurrence of a particular word or where the database contains images without printed text (as in the example of photographs or maps). A full-featured document management system will have user-definable template fields. In situations where the person who selected the keywords is not the one searching for files, this method has obvious limits.

A document management system should allow users to customize index templates, create multiple templates and support different types of index field data within each template, such as date, number and alphanumeric characters. Index fields can be used to categorize documents, track creation or retention dates, or record subject matter, among other information. A document management system should enable pull-down boxes of common key words to speed index field entry and have tools available to help automate entering index information.

Folder/File Structure

Along with enabling full-text and index field searches, a document management system should enable users to locate documents by browsing. A full-featured document management system lets an organization electronically recreate its existing filing system through a nested folder structure. A flexible folder structure eases the transition from paper filing to electronic filing, which makes the transition to document management systems smoother.

4. Distribution, or Putting Information in the Hands of the Right People

A document management system should make it possible for multiple users to access the same files at the same time and for documents to be distributed to authorized individuals within and outside of an organization-over an intranet, by e-mail, or through publication to the Web, CD or DVD. A full-featured document management system safeguards an unalterable copy of the original while allowing you to enhance collaboration and service by circulating copies in the format that best serves your business needs.

When system administrators decide to deploy a document management system across their entire network through an intranet, or even to the public over the Internet, they should make it possible for users to search, retrieve and view documents with any Web browser. Browser-based document access removes the logistical problems associated with computer platform (Windows, Macintosh, Unix, etc.)

5. Security, or the Ability to Protect Your Documents from Loss and Tampering

System security is an absolute necessity for any document management system. A rigorous security system should permit every authorized person to perform required duties – whether from desktop, laptop, the office, a remote location or over the Web – without compromising the integrity of the database, system or network.

A full-featured document management system gives the system administrator the tools to balance access and security through control over both access rights and feature rights. Access rights determine who can log on to the system and which folders or files individuals can open. Feature rights determine the actions that individuals can perform on documents to which they have access. A comprehensive security system also allows high-level users to redact or black out confidential information within files.

Records Management, or Documenting Business Activity

Records management is a specialized branch of document management that deals with information serving as evidence of an organization's business activities. In particular, it is a set of recognized practices related to the life cycle of that information. Most often, records refer to documents, but they can include other forms of information, such as photographs, blueprints, or even books. Records management requires the application of systematic controls to the creation, maintenance and destruction of an organization's records.

The fundamental concept behind records management is the idea of the life cycle of the record. Life cycle refers to the stages that every official business record must undergo. After a record has been created, it must be filed according to a defined, logical scheme into a managed repository where it will be available for retrieval by authorized users. When the information contained in records no longer has any immediate value, the record is removed from active accessibility. Depending on the nature of the record, it is either retained, transferred, archived, or destroyed.

Records management applications should facilitate the inventory of records and the application of consistent records policies. Records management applications must protect records from loss and tampering, while allowing the records manager and other decision makers access to necessary information.

DoD 5015.2 Standard

The Department of Defense (DoD) 5015.2 standard represents the mandatory minimum functional requirements for records management applications used by the Department of Defense agency. While records management applications that have been certified as DoD 5015.2 compliant represent an objective, third-party evaluation, they do not guarantee regulatory compliance or records security. Details regarding this standard are further described in the next section, under records management systems.

III. Essential Components of Document Management Solutions

Although all document management systems provide the basics of scanning, retrieval and display, when it comes to implementing a document management solution in the real-world, system essentials extend far beyond the minimum basics. Document management systems designed for multiple users, a high volume of documents, or multiple office locations must meet more stringent requirements. This section explains what to look for when selecting a document management system for your organization.

Usability

One of the most important factors in how successful a document management system will be is its ease of use. Usability is critical in encouraging fast staff acceptance. A system will only be widely used if it is simple to capture documents, organize and find them. The best systems are user-friendly and flexible enough to adapt to the way people already work within an organization, rather than forcing them to change preferred procedures.

Interface design

To guarantee that a document management system is readily adapted by users throughout an organization, it is important that the graphic interfaces for common operations, such as search and retrieval, be clear easy to use. User-friendly interfaces not only assure rapid adaptation of a document management system by staff, they reduce training expenses associated with implementation.

Capture

For a document management system to enhance business operations, it must accommodate all the types of documents – paper, electronic, fax, audio, video, etc. – that are part of an organization's processes and procedures. It should also enable the batch processing of documents and forms in instances where high-volume processing is part of business operations.

Batch Processing

Organizations that image a significant number of files a day will quickly realize the importance of batch processing. When large numbers of documents need to be brought into the document management system daily, it is inefficient to process each one individually. A full-featured document management system allows files and records to be brought into the system in one batch to speed up the process.

Once all the pages have been captured, the system should let users easily group them into appropriate documents before assigning index fields and moving them to their appropriate folder location. The system should make it possible for pages to be rearranged, removed or added to a document to correct any mistakes that may have occurred in the organization of a file. Similarly, it should be simple to update or add index fields at a later time.

Bar Codes

In high-volume scanning operations, automatically separating and indexing documents using bar codes saves time and money. Bar codes index documents by extracting fields from an external database, by filling in fields with preassigned values, or by associating certain documents with a particular index template. Bar codes can act as markers to indicate the beginning of a new document, automating document separation. While bar codes require some preparation, their benefits can be enormous. For example, if 2000 voter registrations, 500 inquiries and 2500 pages of legislative minutes were to be scanned, bar code stickers could be placed on each document. The system would then automatically read the stickers, determine the start of each new document, assign the correct type of index template for each and fill in template information.

Zone OCR

Organizations that repeatedly process the same forms may want to use Zone OCR (Optical Character Recognition) to reduce data entry time and demands on system memory. Zone OCR saves time through automated document indexing that reads certain regions (zones) of a document and then places information into the appropriate index template fields. The amount of required storage space is also reduced because OCR and indexing are applied only to responses that have been entered.

To minimize errors, the system should allow the user to set a minimum percent accuracy level for OCR. If any portion of the form does not meet this standard, the system should notify the user so that a staff member can read the form and manually enter the correct field information.

Distributed Capture

For organizations with multiple offices, it is important to ensure that a document management system permit users at both central and branch offices to capture and access documents as necessary. Full-featured systems allow for documents to be scanned into the system and transferred into the database at different times so as to minimize traffic demands on the network during peak business hours.

Annotations

Annotations permit users to append or remove information about a document that has been captured without permanently changing the original image. Highlighting, stamps, redactions (black-outs or whiteouts) and sticky notes are among the most common annotations. A document management system's security should give the system administrator control over who can view annotations and see through redactions.

In order for the document to maintain its integrity, all annotations should be overlays that do not change the actual image. This way, a document can be printed with or without the annotations. Although the legal standing of imaged documents varies from state to state, for a document stored in the system to stand up as the best copy of a record, users must not be able to modify the original image.

Storage and Archiving

Non-Proprietary File Formats

Concerns about future readability of documents and records make many organizations hesitant to implement a document management system. With rapid changes in the technology sector, it is hard to predict what applications and hardware will be current five or ten years from now. However, the need for faster retrieval and improved records management means that most organizations cannot wait to implement solutions.

To address these concerns, document management systems should use non-proprietary image and text formats. As the example of word processing makes clear, documents created and saved with obsolete versions of a program can be difficult or even impossible to read. Since each word processing software company uses proprietary formats for its documents, converting files from old versions can be a frustrating or expensive task. Similar compatibility issues apply to the document management world.

The non-proprietary formats available for storing document information are few, but stable. ASCII has been a standard for text information since 1963 and is a basic building block for practically every text-based program. TIFF has been used as a standard, non-proprietary graphics format since 1981. It is widely used to transmit document information by document management systems, fax machines and other software. Given the prevalence of ASCII and TIFF, system purchasers can feel confident that no matter what new paradigm arises in the future, the developers of the new format will have a vested interest in providing a conversion for these standards. With proprietary document formats, there is no such assurance.

Portability and CDs/DVDs

Document management systems should enable users to carry important documents anywhere for convenient viewing on other computers. When people go on business trips, they often need to bring key documents with them. Carrying paper documents is often impractical, and copying an entire database to a laptop can be impossible. With a document management system that supports briefcases or portable volumes, documents can be detached or copied and moved to other databases in other locations. Document management folders containing relevant documents can be transferred to other databases quickly and easily using searchable CDs that hold up to 650 MBs of data – the equivalent of approximately 12,000 pages – or read-only DVDs, whose capacity ranges from 4.7 to 17.1 GBs.

If a document management system does not provide this level of document portability, users of the system will find it difficult to bring their documents on the road and to transfer files between different offices. Briefcases and portable volumes help users to transfer their documents to other offices, laptops or customers quickly and easily. Optical discs also weigh much less than paper files.

Briefcases

A full-featured document management system allows users to simply drag and drop the appropriate document management system folders into a briefcase and transfer the briefcase to a laptop computer or a computer in a remote office.

Portable Volumes

Portable volumes are high-volume briefcases that allow for constant updates to shared document management databases in different locations. This ability is useful for organizations that use a scanning bureau on an ongoing basis or for organizations with multiple offices. On many large-scale document management systems, the document files are stored on multiple drives or network volumes. Portable volumes allow entire volumes containing document images and text to be transferred en masse to another database.

Indexing and Retrieval

For a document management system to support multiple users with different job functions, it is essential that it enable multiple means of searching for information.

Types of Searches

There are several helpful options to maximize the effectiveness of full-text searches.

Fuzzy Logic

Most searches assume that the search words have been spelled correctly and perfectly indexed by the OCR process or during the manual entry of index fields. Unfortunately, people frequently misspell words, and no OCR process is 100% perfect. Fuzzy logic compensates for these errors by searching for spelling variations. A document management system should allow the user to control the search by setting how many letters can be wrong or what percentage of a word can be wrong. For example, a fuzzy logic search for "goat" would find "goat," "gout" and "coat."

Wildcards

Wildcards are characters, like the asterisk (*) and the question mark, which can be used in searches to compensate for misspellings or unknown spelling. The asterisk stands for any character or characters, while the question mark stands for any single character. For example, searching for "c*t" would find the words "cat," "cot," "coat," "cut" and "chest." Searching for "c?t" would only find the words "cat," "cot" and "cut."

Boolean Operators

Whenever full-text searches are performed, there are usually several documents that meet the search criteria. Boolean operators (AND, OR and NOT) help fine-tune searches and reduce the number of unrelated documents on the results page. For example, to find documents relating to the former governor of California and not to the University of California at Davis, users could search for "Davis AND governor."

Proximity Searches

Proximity searches can also be used to narrow the search results. They are used to find words that occur within a certain number of words, sentences or paragraphs of each other. For example, to find documents relating to tobacco lawsuits, but not smoking ordinances or tobacco growing, users could search for "tobacco" within one sentence of the word "lawsuit."

Result Display

The way in which the results of searches are displayed has considerable impact on the usability of a document management system.

Lines of Context

Even the most specific full-text searches can produce several hits when large document databases are involved. In addition to providing users with a list of documents that meet their search criteria, some document management systems reveal lines of context that display each occurrence of the search word in each document. Lines of context help users pinpoint the appropriate document without having to view every document in the search results.

Highlighted Search Words

Once a document is identified, the search word needs to be located within the document. To help with this, some document management systems display the appropriate page of the document and highlight the search word in both the text and on the document image. This makes it easy for the user to immediately zoom in on the relevant section instead of having to look through multiple pages of a document. The importance of this becomes obvious when the needed word occurs on page 97 in a 200-page document.

Distribution

Document management systems must provide efficient ways of getting information out of the system on the level of the individual document. Printing, faxing and e-mailing documents are several ways of doing this. Document management systems should promote the rapid copying of files to a CD or DVD. To be most effective, the document management system should support royalty-free CD or DVD duplication and contain viewers that enable people without a document management system to search for and view documents on the disc.

Print/Fax/E-mail

To maximize their usefulness, document management systems should support the most common printer and fax drivers and be able to print images, text and annotations.

E-mail has become the default mode of communication in many organizations. Organizations obtain significant gains in efficiency and save considerable expense by transmitting documents via e-mail instead of using faxes, courier services or the postal service. Document management systems should have options that make it possible for images to be easily sent with any MAPI (Mail Application Program Interface)-compliant e-mail system and read by recipients who do not have document management systems.

Internet/intranet

A document management system should provide a simple way to publish information to the Internet or an intranet. This allows organizations to share information with other departments, remote offices, clients or the public. Web systems should be fully searchable and must support the same security protocols as network systems. Ideally, a document management system will require no HTML or complex coding to post files to the Web.

Workflow

Workflow can increase the benefits of a document management system by automating the routing of documents to various people, which eliminates bottlenecks and streamlines business processes. This added functionality is more important for large offices, for organizations with central and branch offices and for organizations that plant to expand their system.

Workflow should automatically notify specific users of specific document-management-related system events, based on rules created by the system administrator. Workflow should generate return receipts and timed responses. If a recipient does not act within a specific time frame, the program should send either a reminder message or a second message to an alternate recipient. An essential component in any procedural workflow system is document automation. Workflow should be able to automatically move, copy or delete documents within the document management database based on a predetermined set of rules. However, the success of any workflow system is not its ability to follow the strict routing and reporting features of a fully automated system, but its ability to handle exceptions to the rules as they arise. An effective workflow system provides the system administrator complete access to on-the-fly routing of documents and information through the system's folder structure and system security.

Workflow systems should offer administrators drag and drop simplicity, an intuitive graphical interface and an easily understood folder structure. Workflow applications should be ODBC-compliant (see Integration, below) to facilitate integration and customized applications. As a final component, workflow must provide for comprehensive security reporting through an audit trail function.

Security

Security is critical to the successful implementation and ongoing protection of a document management system. While security may not be the primary concern for a single department installation, it becomes more important as the system is expanded to allow different departments, and even the public, access to files. A document management system should provide multiple levels of security including authentication, authorization, audit trails, and disaster planning. The system's security should parallel that of the network and be simple to administer.

Authentication

Authentication is the level of security that requires users to present credentials, normally a user name and password, in order to access the system.

Authorization

Authorization is the level of security that controls access to objects (files, folders, etc.). Authorization encompasses two primary areas: access rights, which determine the objects users can open, and feature rights, which determine the actions that users can perform on the objects to which they have access.

Access Rights

A document management system should let organizations assign access to specific folders, as well as specific documents, at both the group and individual level. The use of groups with inherited, or predefined, rights allows system administrators to quickly assign viewing privileges, while individual-level security allows specific users such as managers to also view documents that the rest of the group cannot. For example, access rights would allow the system administrator to deny most employees in an organization access to HR files, while allowing human resource staff members to view the personnel files of everyone in the organization except other HR personnel and the HR director to view all personnel files.

A full-featured document management system will not allow users to see objects for which they do not have viewing privileges. This protective feature is especially important for organizations whose systems contain confidential files and folders.

Feature Rights

A document management system should also let system administrators limit the actions that users are entitled to perform on folders and documents at both the individual and group level. Feature rights determine a range of actions, including adding pages, annotating, copying, or deleting records. For example, a system administrator could allow various departments to have viewing privileges to city council minutes, but allow only the City Clerk to have annotation rights to those files.

Redaction

Redaction (blackout or whiteout) is a security feature applied within documents to make certain portions of the document inaccessible, except to authorized users. A document management system should offer the ability to redact portions of a document's image and/or text. Users' ability to view redacted text would depend on their security rights. For example, a system administrator could make crime reports available to various city departments, but allow only the Police Department to see sensitive information such as the victim's name and address.

Audit Trails and Reporting

As an additional level of security, a document management system should offer the ability to generate audit trails and reports that detail system activity. A document management system should be able to log all users, documents viewed, actions performed and the time at which the actions were performed. A full-featured document management system will log unsuccessful attempts to perform actions and provide electronic watermarks to authenticate printed documents. Audit trail abilities are especially important when an organization has many different users and confidential documents. Audit trails also play significant roles in demonstrating regulatory compliance.

Disaster Recovery

Digital archiving with a document management system simplifies disaster recovery planning by allowing backups of entire document repositories to be stored on durable CDs, DVDs or other media. In the event of document loss or damage, archives can be reconstructed from digital backups. The solution should also provide built-in viewers on published CDs and DVDs. This allows immediate document access from any PC with a CD or DVD drive even if the network remains offline for an extended period.

Integration

The introduction of new software and databases often creates logistical challenges for the computer support staff of an organization. Document management programs should offer packaged integration tools for simple image enabling in order to minimize the burden on computer support staff. To minimize disruptions to business operations, it is essential that a document management system integrate smoothly with other software applications, such as PeopleSoft, Geographic Information Systems (GIS) and Student Information Systems (SIS), and databases in use by the organization.

Back-end characteristics that facilitate integration include:

- Open architecture – the use of hardware and software whose specifications are designed for easy integration. This enables anyone to create add-on products to connect the hardware or software to other devices.
- UDA (Universal Data Access) compliance – conformity to a single application program interface designed by Microsoft that allows users to combine data from different databases made by various manufacturers.

See also System Compatibility below.

Technical Considerations

System Compatibility

Compatibility is the capacity of a document management system to work with existing hardware and software systems. To maximize compatibility with your existing systems, a document management system should:

- Work with standard operating systems and support standard database platforms.
- Communicate using popular network protocols such as IPX/SPX or TCP/IP.
- Have the capability to deploy over the Web.
- Use n-tier architecture with client-side image compression/decompression and server-side searching and indexing to minimize network traffic.
- Store text and image files in non-proprietar industry-standard formats.

Networked Systems

In any office, documents are used to transmit information between people. For document management to be truly useful in an office environment, documents must be accessible to all authorized users. It is important for document management systems to have a central repository of records, accessible from any PC. Storing documents on individual PCs, however, impairs the flow of information between coworkers and wastes valuable time and resources. Networked systems are vital to foster collaboration. Networked systems carry out certain document management functions more efficiently than individual PCs can. For example, Optical Character Recognition (OCR) of an image requires a great deal of computing power.

Scalability

The scalability of a system determines how much it can grow with your organization's needs. For full scalability, a system should:

- Support the entire group of users within an organization concurrently.
- Store all documents in the organization.
- Be designed to accommodate a high volume of users and documents.
- Store information across multiple drives or servers.
- Support multiple databases.
- Accommodate high-volume usage.
- Integrate with other applications.

n-Tier Architecture

Document management solutions, like any other network application, consume computer resources. Image files are large, and databases must track large numbers of records. Functions such as OCR, image display and searching require extensive computing power. It is important to have n-tier architecture when more than a few people need access to imaged documents. Even when an installation begins with a single-user pilot project, it remains important that the document management system be able to accommodate future growth.

An n-tier system delivers maximum scalability in departmental solutions and across the organization with distinct client, business logic, data and document layers. Any network-connected storage media, including Storage Area Networks (SANs), can be used for physical storage, while multiple SQL servers handle the distributed database layer.

Tasks such as indexing, OCR and searching are distributed between the client (PC workstation) and the document management server for optimal performance. Some tasks are performed more efficiently on the client, while others are better handled by the central server. Where the specific tasks are performed may vary among different document management systems.

It is important to distinguish between this more robust design approach and simple file-sharing applications. In file-sharing applications, file integrity can be compromised when a workstation program is interrupted in the middle of a transaction. With computing functions distributed across multiple tiers, however, the client does not open data files directly. Therefore client interruptions do not threaten data integrity.

An n-tier system can perform searches much faster, since the server machine is typically more powerful than individual workstations. File-sharing systems send a copy of the entire database over the network to the workstation, which then performs the search. This method leads to: a) increased network traffic if, for example, the database is 800 MB in size; and b) search response times that are dependent upon the speed of the PC workstation. File-sharing systems may be easier to develop and therefore less expensive initially, but their design ultimately restricts flexibility and scalability, limiting their long-term usefulness.

Thin Client

A thin client is an infrastructure-friendly solution that minimizes the burden of application installation, maintenance and software upgrades. The benefits of thin clients extend beyond conserving IT resources to expediting the search and retrieval of information over the organization's intranet or the Web. A Web-browser-based thin client must effectively deliver essential features to end users without compromising system security.

Records Management Applications

Records management systems enable the application of systematic controls and policies concerning the life cycle of those records that detail an organization's business transactions. Records management applications (RMAs) should allow organizations to file records according to a determined scheme, to control the life cycle of records, to retrieve records based on partial information and to identify records that are due for final disposition.

Records Series and Metadata

A records management application must allow records to be refiled in different folders or series after their initial filing in order to meet DoD 5015.2 criteria. A records management application must also have a way to control the metadata fields associated with every record, record series and record folder. It must limit the entering of metadata to the time of filing, yet allow authorized users to edit and correct filing errors.

Linking and Versioning

The records management application must allow users to indicate related records through linking, a form of metadata that defines and establishes relationships between documents. Examples include supporting documents, superceded/successor records, multiple renditions and incremented versioning. A records management application should allow document links to be established by all users at the time of filing, but only authorized users should be able to create, modify or remove links post-filing.

Versioning is a special document relationship, used to indicate an auto-incremented sequence of revisions to a particular record. The records management application must allow users to establish record versioning. Versions must be retrievable as if they were independent documents and contained their own metadata. A records management application must clearly indicate if a record has multiple versions and which version is the most recent.

Security Tags and Audits

Security tags represent a metadata field intended to define and restrict access to records, as well as aid in their classification and retrieval. A records management application must allow the records manager to define security tags and to allow users to assign tags to records upon filing. Only authorized users should be able to modify or remove security tags post-filing. The records management application must also support the audit of all filing, handling and disposition of records.

Vital Records

Vital records – those records deemed essential in order for an organization to resume business operations immediately after a disaster – are subject to periodic review and update. A records management application must provide a way to assign a review cycle to vital records and detail when they were last reviewed. Examples of vital records include emergency operating records or legal and financial rights records. The records management application must also offer a way to retrieve all vital records, identify when they were last reviewed and indicate vital records due for review at any given moment.

Disposition and Freezing

The records management application must handle two types of disposition action: interim transfers and final disposition. The available actions for final disposition are accession and destruction. The records management application must allow for the exportation of entire record folders and their metadata values for transfer and accession events. Following the confirmation of successful transfer, the records management application should be able to maintain the records, maintain only the metadata or completely delete the records. The records management application should be able to freeze a folder. When a folder is frozen, no record may be removed from the folder, and no record in the folder may be modified.

IV. Implementation: Addressing Your Business Needs

When you consider document management systems, there are a number of factors to keep in mind.

- How many documents must the system store? Consider both the number of existing documents and the number of documents added annually. This information detemines how much storage space is needed, the hardware configuration and the cost of the system.

- How many users will be using the system concurrently? This determines preliminary software costs, required licenses and server size.

- What departments will be using the system and will it be necessary to provide public access? This determines what specific features and levels of security will be needed.

- What business problems need to be solved to reduce costs and improve productivity? This determines which functions of a document management system will be requirements and which will be optional. It also helps determine whether plug-ins or customizations will be needed.

- Are there regulatory compliance issues governing your organization? If so the document management system should have functions that support compliance.

- Do you need to integrate your document management program with other software applications, such as human resource or GIS (Geographic Information Systems) programs? Because integration issues often increase the time required for implementation of document and records management systems, these concerns should be resolved before investing in a particular system.

- Do you want a turnkey solution or a customized one? This determines the amount of consulting, installation, training, configuration and support that will be needed.

- What type of network is currently used and will it continue to stay in place, or will it be upgraded? This determines net work constraints, system configuration and workstation upgrades.

Records Management Considerations

Records management systems necessitate special considerations in addition to those listed above.

- The records management application should support custom searches based on record properties, retention or disposition properties, full-text content, template fields, folder location, sticky-note contents and more. It should be possible to save search results in a usable format, such as an Excel spreadsheet.

- The records management application should manage the full life cycle of the record, from document creation through declaration as a record to final disposition.

Scaling from Pilot Project to an Organization-wide Solution

Large organizations sometimes prefer to begin with a pilot project involving one or two departments before expanding their document management system to the entire organization. Whether or not an organization begins with a pilot project, a document management system should be scalable, meaning that it should allow an organization to easily expand the size of the system to accommodate organizational growth, at the level of either users or documents.

Installation

The first step of an installation should be a site evaluation by the software vendor to determine proper equipment placement and to identify any network connectivity problems. Hardware installation consists of connecting and setting up all components, including installation of the necessary operating systems and drivers. It requires the testing of equipment to ensure proper hardware functionality and network connectivity.

After tests of the hardware have been conducted, the document management software is installed on the document management server and the necessary workstations. It must be tested to ensure operability. Generally, the software vendor will perform these tasks with the collaboration of the organization's IT personnel.

Training

Training programs should be tailored to the specific needs of different levels of users and their concerns.

End User

End-user training involves a focus on the basics of daily system use. This training should take place on-site. Each group should receive all instruction necessary to ensure comfort with the new document management system. The amount of training necessary will depend on the users' level of familiarity with Windows applications, the document management system's ease of use and the degree of change from existing procedures. Because of the need to bring new employees up to speed as quickly as possible, a well-designed document management system should be easy to use.

Given a user-friendly system and minimal change in procedures, most users will become proficient in a short time period. Effectiveness is improved when the class size is limited to no more than 10 people and participants are free from interruption. Training should include supervised, hands-on use of the document management system during actual operation. This allows users to ask questions that might not occur to them until they are using the system for business procedures.

System Administration

It is important to train select individuals on how to administer and maintain the system. On-site training is recommended because it increases familiarity with specific details of the document management system.

Implementation Consulting

Implementation consulting assists those responsible for the document and records management functions to develop strategies for translating the organization's current filing and indexing structures into electronic systems. Electronic filing is different from paper filing, and records managers face the challenge of these differences when setting up their systems. Considerations regarding retention schedules, storage and filing methodologies need to be evaluated before the system is fully implemented. The length of the training depends on the complexity of the filing system and should take place on-site.

Support and Maintenance

Document management systems, like any mechanical tool, require maintenance. Organizations should evaluate the software vendor's support structure. Vendors should offer various levels of support from software upgrades to regular, on-site maintenance visits.

Factors that affect the level of support that an organization needs are:

- Size of the system purchased
- Amount of time demands on the system
- IT personnel's level of experience with document management
- Internet access
- Concurrent changes that have to be made to the organization's computer network or infrastructure
- Rate of personnel turnover

Support can entail any or all of the following:

- Software upgrades
- Telephone hotline support
- Online forums
- Remote dial-in access to your system
- Software patches available through an FTP site
- Regularly published technical bulletins or newsletters
- On-site maintenance visits
- Additional and/or advanced training sessions
- Hardware support

When purchasing hardware, such as servers, storage devices and workstations, organizations should choose vendors with good reputations for service and support. While the initial cost may be higher, the benefits include less downtime and more consistent, reliable functioning.

Outsourcing Scanning

Organizations sometimes find it faster or more cost effective to have a service bureau perform their back-file document conversion or ongoing document scanning. Generally, in these cases, the document management system is maintained by the organization, while the service bureau is responsible for delivery of the scanned documents on CDs or another medium. In addition to storing images and text information, these CDs must also carry data describing the document names, index fields, folders, etc.

If the organization has been modifying existing documents and creating new ones during this time, overwriting the organization's database with the new one provided by the service bureau is not an option. The document management system must be able to merge new and existing data seamlessly. A portable volumes feature will handle this automatically.

Compliance and Legal Issues

A document management system can help limit exposure to civil and criminal liability stemming from non-compliance with regulatory statutes by ensuring the consistent application of policies organization-wide and by providing audit reports.

While laws and auditing authorities vary by industry and state or region, most regulations share two common principles: the information must be set in time, meaning that the date and the time of the creation of the digital images must be recorded in an unalterable fashion, and the storage media used by the system must be unalterable. In some areas, such as financial planning, a copy of the records must be maintained by an independent third-party and be readily available to auditors, when requested.

In order to meet general compliance demands, a document management system must:

- Allow documents and records to be retrieved on demand.
- Store digital images on acceptable media.
- Maintain records in an unalterable format.
- Permit a complete and accurate transfer of records.
- Possess reasonable controls to ensure integrity, accuracy and reliability.
- Have reasonable controls to prevent and detect the unauthorized creation, alteration or deletion of records, as well as record deterioration.
- Contain an indexing system that facilitates document retrieval.
- Be able to print copies of records, when required.
- Make cross-referencing with other record-keeping systems and software possible.
- Have documentation on how the software works and how it is set up.

Many government agencies now accept imaged documents as legal records, meaning that the paper originals can be destroyed, given certain conditions.

In general, for an imaged document to qualify as a legal record, the following must be true:

- Records must be stored in an unalterable format, such as CD, DVD or WORM.
- The system must have controls to ensure integrity, accuracy and reliability.
- The system must provide some type of audit trail to prevent and detect unauthorized creation of, addition to, alteration of or deletion of records.
- A complete and accurate transfer of records must be possible.
- The system must have reasonable controls to prevent and detect deterioration of records.
- There must be an indexing system to assist with finding records.
- The system must have the ability to print copies of records.
- The system must be able to cross-reference other record-keeping systems and software.
- The system must have documentation on how the software works and how it has been set up.

The legality of imaged documents varies depending upon the federal agency, state, county, municipality and department involved. Organizations should consult with an attorney on the specific statutes governing their area.

V. Frequently Asked Questions

General

Q. What is a document?

A. A document consists of information stored on anywhere from one to several thousand pages. It can include images and/or text, plus annotations, and one template (index card).

Q. Can I edit or alter images?

A. A document management system should not allow the original image to be altered or edited. Annotations should be overlays that do not alter the original document. It is important to protect the original image in order to maintain both the legal status of the document and the integrity of the system.

Q. Do document management systems support audit trails?

A. Yes. A document management system's audit trail should record username, date, time, document name and action for every instance in which a user accesses a database or document. Various levels of audit trail logging detail and activity tracking should be available. The system should include a viewer to sort and filter these logs. Audit trails are especially important for regulatory compliance.

Q. What is the standard format used to store images?

A. Black and white images are most commonly stored as standard TIFF files using CCITT Group IV (two-dimensional) compression. Grayscale and color images are frequently stored as TIFF files with JPEG compression.

Q. What is the standard format used to store text?

A. ASCII, which stands for the American Standard Code for Information Interchange, has been the standard, non-proprietary text format since 1963.

Q. How much disk space does a document management system typically require?

A. A single page typically occupies around 50KB of disk space, if the image is stored in TIFF Group IV. Each gigabyte (GB) of storage space, which amounts to only a few dollars, holds approximately 20,000 pages. With the significant drop in prices for hard drives and optical media, it costs much less to store documents in a document management system than on paper.

Q. What if my database is too big to fit in one data volume?

A. A document management system should allow data and images to be stored across multiple volumes, with each volume residing in a different directory or on a different drive, disk array, CD or MO disk.

Capture

Q. What are the most common hardware and software scanner interfaces?

A. Many scanners attach to an Adaptec SCSI card or to a Kofax Image processing board. Most scanners use either TWAIN or ISIS scanner drivers to communicate with the computer.

Q. How can I scan forms?

A. Forms processing components often use multiple OCR engines and elaborate data validation routines to extract hand-written or poor-quality print from forms that go into a database. Because many forms that are scanned were never designed for imaging or OCR, it is essential to have good quality assurance mechanisms in place when scanning forms to correct errors that might occur.

Q. Can I capture information from multi-function peripherals (MFPs)?

A. A full-featured document management system allows you to capture documents from different network locations, including MFPs, or devices that perform any combination of scanning, printing, faxing or copying.

Q. How can I scan large format documents?

A. Several manufacturers make scanners specifically designed for large format documents up to E-size (34 inches x 44 inches) and A-0 size (33 inches x 46.8 iches). If you do not have one of these, the document can be reduced in size using a photocopier and then scanned with a normal scanner, or sent to a service bureau that has large format scanners.

Q. What image resolution should I use?

A. Most imaging systems can support documents scanned at various resolutions, from 50 dpi to 600 dpi (or more) depending on your scanner. Depending on the purpose and the contents of the page, most documents are scanned in black and white at 300 dpi.

Q. What about color files or photographs?

A. Imaging systems should support black and white, grayscale and color images. Color files can be scanned with a color scanner or imported into a document management system. There are a wide range of color scanners on the market. Many document management scanners support color and grayscale.

Q. How can I scan double-sided documents?

A. An imaging system should provide two different ways to do this. It should support duplex scanners, which simultaneously scan both sides of a page, and simplex scanners, which require the user to scan all the front sides, place the documents in upside down and then scan all the back sides, before the system collates the pages into the correct order.

Q. Can I scan landscape and portrait pages together?

A. An imaging system should allow you to change the orientation of pages during or after scanning. A well-designed system will also include an option to automatically check and correct the orientation of pages.

Q. How are skewed images handled?

A. Skewed (crooked or tilted) images can adversely affect the accuracy of the OCR process, so an imaging system should include software that recognizes skewed images and compensates for them. This is particularly important when scanning press cuttings on a flat bed scanner or when scanning documents through a worn-out or poorly designed automatic document feeder (ADF).

Q. How can I scan checks?

A. Several manufacturers make scanners specifically designed for checks, which read the magnetically encoded MICR (Magnetic Ink Character Recognition) numbers at the bottom of the check. If you do not have one of these scanners, most checks can be scanned with regular document imaging scanners and OCR-processed as usual, though the MICR numbers will not be read. To integrate MICR information into the document management database, the document management system must support check scanning hardware.

Q. What file formats can a versatile system import?

A. A versatile system should be able to import the files you encounter in your office. This includes word processing files, spreadsheets and presentations as well as common image formats such as TIFF Group IV, TIFF Group III, TIFF Raw, TIFF LZW, PCX, BMP, CALS, JPEG, GIF, PICT, PNG and EPS Preview images. A document management system providing long-term archival of documents should allow the images of each page to be stored in a non-proprietary format. For example, electronic document pages would be printed to the document management system, black and white graphical files would be converted to TIFF Group IV format and color/grayscale images would be converted to TIFF or JPEG.

Indexing

Q. How do I index scanned documents?

A. There are three primary ways to index documents: folder structure, index or template fields, and full-text indexing. Folder structure essentially functions as a visual indexing method that allows users to browse for documents by categories. Index or template fields categorize documents according to keywords, which can be either manually entered or automatically assigned by the document management program. Full-text indexing is the automated process of entering every word in a document into the index.

Q. What is OCR?

A. OCR stands for Optical Character Recognition and refers to the way a computer converts words from an unsearchable scanned image to searchable text. OCR is usually necessary in order to use full-text indexing and searches, so it should be included in an imaging and document management system. OCR engines can generally only recognize typed or laser-printed text, not handwriting.

Q. What is the difference between OCR and indexing?

A. OCR is the process of converting scanned images to text files. Full-text indexing is the process of adding each word from a text file to an index that specifies the location of every word on every document. Well-designed document management software can make this a fast and easy procedure, providing rapid access to any word in any document.

Q: What is the difference between index field searches and full-text searches?

A: Index field or template searches enable you to retrieve preestablished categories of documents, whereas full-text searches turn up every occurrence of designated words in the database. When the database contains a large number of documents, the difference between sorting documents by topic and listing every occurrence of a word in the database – including passing references – is significant in terms of the time required to analyze the search results and locate the desired document(s).

Q. How accurate is OCR?

A. Accuracy on a freshly laser-printed page is typically better than 99.6%. Accuracy on faxed, dirty or degraded documents will be lower, so it is essential that an imaging system have image clean-up technology to improve OCR accuracy.

Q. Do I have to go through text to correct OCR mistakes manually?

A. Well-designed systems allow users to correct OCR errors from within the system. However, when hundreds or thousands of pages are scanned every day, it is usually not practical to clean up the text. Because the OCR process does not have perfect accuracy, it is important that the document management system support fuzzy logic searches. Fuzzy logic searches allow for misspelling and will find words even if the OCR engine makes occasional mistakes.

Q. How fast is the OCR process?

A. The performance of the OCR and indexing processes is entirely dependent on factors such as the speed and configuration of the host system as well as the contents of the image.

Q. What is ICR (Intelligent Character Recognition)?

A. ICR is pattern-based character recognition and is also known as Hand-Print Recognition. Handwritten text is more difficult for computers to recognize and results in higher error rates than printed text. ICR engines usually do best at recognizing constrained printing, which means block printed letters with one letter in each box. Accurate recognition of unconstrained handwriting, especially cursive handwriting, typically requires that the ICR engine be trained to recognize each user's style of writing.

Q. What is OMR (Optical Mark Recognition)?

A. OMR, also called Mark-Sense Recognition, is the recognition of marks commonly used on forms, such as check marks, circled choices and filled-in bubbles. OMR can be an important part of a document management system for organizations that process many standard forms. Exam forms and customer survey cards are perhaps the best-known examples of OMR.

Q. Can OCR-processed text be exported and reused in a word processor?

A. Yes, you can usually cut and paste text between the document management system and another Windows application, or you can export complete text files (all text pages in a document) to a directory and open it with your preferred word processing program.

Viewing/Printing/Exporting

Q. Can I open and display more than one document at a time?

A. Some document management systems will allow you to display multiple documents, with the number of documents that you can have open simultaneously limited only by the amount of memory available.

Q. How can I resequence pages of a document before printing or exporting?

A. If pages are out of order and need to be resequenced, a well-designed document management system will allow you to drag thumbnail views of pages to the required position. In the same way, individual pages can be selected and deleted, subject to appropriate security access control and privileges.

Q. What is the advantage of a large monitor?

A. For people who use an imaging system frequently, screen size can be a critical factor. If users are to flip between pages with the ease of real paper, they must be able to view the whole page at once in a way that allows the text to be readable. If 8$^1/_2$-inch x 11-inch pages are the dominant paper size, then a 21-inch monitor capable of displaying 1600 x 1200 is optimal. Using a 15-inch VGA monitor will require scrolling and panning if the image is viewed at normal size.

Q. What other display considerations are important?

A. Screen resolution and the refresh rate of the monitor are also important. Generally, the larger a monitor is and the higher resolution it has, the harder it is to get the high refresh rate that is required for sustained viewing without screen flicker. The optimum threshold for minimum flicker is generally considered to be a horizontal refresh rate of 72 MHz on a 21-inch monitor. The maximum refresh rate is a function of the monitor and the graphics controller.

Q. Will I need a specialized printer for images or OCR-processed text?

A. Generally no. Most imaging systems support a wide variety of Windows-compatible printers, but an optimal configuration includes a laser printer with at least 4 MB of RAM. If you are using a networked system and printing high volumes of pages to a network printer, you might consider installing a separate laser printer either locally or on its own network segment to minimize network traffic.

Q. In which formats can I export documents?

A. It depends on the document management system. Common graphical formats include TIFF Group III, TIFF Group IV, TIFF Raw, BMP, PCX, PNG and JPEG.

Q. What happens when a user without redaction viewing rights prints a document that has been redacted?

A. A document management system should protect the integrity of the document by printing with the redactions intact.

Records Management

Q. Are all documents records?

A. No. Records management is a specialized discipline that deals with information serving as evidence of an organization's business activities. In particular, it is a set of recognized practices related to the life cycle of that information. Often, records refer to documents, but they can include other forms of information, such as photographs, blueprints or even books.

Q. What does records management software do?

A. Records management software supports the application of systematic controls to the creation, maintenance and destruction of an organization's records.

Q: Does DoD 5015.2 certification guarantee compliance with other regulations like HIPAA?

A. No. It is important to distinguish between regulatory compliance and the DoD 5015.2 standard. The DoD standard represents baseline functionality for records management applications (RMAs) used within the Department of Defense. It serves as the de facto standard for records management applications across government and industry. However, it is a records management standard and not a broad regulatory compliance standard. DoD-5015.2 certification facilitates compliance by supporting the application of systematic records policies; it cannot guarantee compliance. Compliance is a process dependent on the application of records policies.

Q: How do records management applications help enforce proper polcies?

A: Records management applications can support the application of consistent policies and procedures through a series of mechanisms, including: mandatory metadata acquisition and automated extraction of e-mail metadata; support for time, event and time-event dispositions; automated notification for review of vital records; freezing of records; and comprehensive audit trail reporting.

COLD (Computer Output to Laser Disc)

Q. What is the difference between COLD and imaging?

A. COLD is specifically for archiving, indexing, searching and printing reports from high-volume text files generated by mainframes, mini-computers and other computer applications. COLD stores large report files and extracted index fields on hard disk, optical cartridge or CD-ROM instead of printing all the information out on paper or storing it to microfilm.

Q. How many index fields can the COLD server extract from each report?

A. The number of index fields is usually unlimited. However, the more fields extracted from each report, the more slowly the extraction process will run and the larger the index files will be.

VI. Glossary of Terms

Access Rights

A security mechanism that lets the system administrator determine which objects (folders, documents, etc.) users can open. It should be possible to set access rights should for groups and individuals.

ADF

Automatic Document Feeder. This is the means by which a scanner feeds the paper document.

Annotations

The changes or additions made to a document using sticky notes, a highlighter or other electronic tools. Document images or text can be highlighted in different colors, redacted (blacked-out or whited-out) or stamped (e.g., FAXED or CONFIDENTIAL), or have electronic sticky notes attached. Annotations should be overlaid and not alter the original document.

ASCII

American Standard Code for Information Interchange. Used to define computer text that was built on a set of 255 alphanumeric and control characters. ASCII has been a standard, non-proprietary text format since 1963.

ASP (Active Server Pages)

A technology that simplifies customization and integration of Web applications. ASPs reside on a Web server and contain a mixture of HTML code and server-side scripts. An example of ASP usage includes having a server accept a request from a client, perform a query on a database and then return the results of the query in HTML format for viewing by a Web browser.

Audit Trail

An electronic means of tracking all access to a system, document or record, including the modification, deletion and addition of documents and records.

Bar Code

A small pattern of lines read by a laser or an optical scanner, which correspond to a record in a database. An add-on component to document management software, bar-code recognition is designed to increase the speed with which documents can be stored or archived.

Batch Processing

The name of the technique used to input a large amount of information in a single step, as opposed to individual processes.

Bitmap/Bitmapped

See Raster/Rasterized.

BMP

The abbreviation for a native file format of Windows for storing images called bitmaps.

Boolean Logic

The use of the terms AND, OR and NOT in conducting searches. Used to widen or narrow the scope of a search.

Briefcase

A method to simplify the transport of a group of documents from one computer to another.

Burn (CDs or DVDs)

To record or write data on a CD or DVD.

Caching (of Images)

The temporary storage of image files on a hard disk for later migration to permanent storage, like an optical or CD jukebox.

CD or DVD Publishing

An alternative to photocopying large volumes of paper documents. This method involves coupling image and text documents with viewer software on CDs or DVDs. It is essential that search software be included on the CDs or DVDs to provide immediate retrieval abilities.

CD-R

Short for CD-Recordable. A CD that can be written (or burned) only once. It can be copied as a means to distribute a large amount of data. CD-Rs can be read on any CD-ROM drive whether on a standalone computer or network system. This makes interchange between systems easier.

CD-ROM

Compact Disc-Read Only Memory. Written on a large scale and not on a standard computer CD burner (CD writer). An optical disc storage medium popular for storing computer files as well as digitally recorded music.

Client-Server Architecture vs. File-Sharing

Two common application software architectures found on computer networks. With file-sharing applications, all searches occur on the workstation, while the document database resides on the server. With client-server architecture, CPU-intensive processes (such as searching and indexing) are completed on the server, while image viewing occurs on the client. File-sharing applications are easier to develop, but they tend to generate tremendous network data traffic in document management applications. They also expose the database to corruption through workstation interruptions. Client-server applications are more difficult to develop, but dramatically reduce network data traffic and insulate the database from workstation interruptions. See also n-Tier Architecture.

COLD

Computer Output to Laser Disc. A process that outputs electronic records and printed reports to laser disc instead of a printer. Can be used to replace COM (Computer Output to Microfilm) or printed reports such as greenbar.

Compression Ratio

The ratio of the file sizes of a compressed file to an uncompressed file. With a 20-to-1 compression ratio, an uncompressed file of 1 MB is compressed to 50 KB.

Deshading

Removing shaded areas to render images more easily recognizable by OCR.

Deskewing

The process of straightening skewed (off-center) images. Documents can become skewed when they are scanned or faxed. Deskewing is one of the image enhancements that can improve OCR accuracy.

Despeckling

Removing isolated speckles from an image file. Speckles can develop when a document is scanned or faxed.

Disposition

Actions taken regarding records after they are no longer required to conduct current business. Possible actions include transfer, archiving and destruction.

Dithering

The process of converting grays to different densities of black dots, usually for the purposes of printing or storing color or grayscale images as black and white images.

Document Management

Software used to store, manage, retrieve and distribute digital and electronic documents, as well as scanned paper documents.

Duplex Scanners vs. Double-Sided Scanning

Duplex scanners automatically scan both sides of a double-sided page, producing two images at once. Double-sided scanning uses a single-sided scanner to scan both pages, scanning one collated stack of paper, then flipping it over and scanning the other side.

DVD

Digital Video Disc or Digital Versatile Disc. A disc similar to a CD, on which data can be written and read. DVDs are faster, hold more information and support more data formats than CDs.

Feature Rights

A security mechanism that allows system administrators to determine the actions that users can perform on the objects to which they have access.

Flatbed Scanner

A flat-surface scanner that allows users to capture pages of bound books and other non-standard-format documents.

Folder Browser

A system of on-screen folders (usually represented as hierarchical, or stacked) used to organize documents. For example, the Windows Explorer program in Microsoft¨ Windows is a type of folder browser that displays the directories on your disk.

Forms Processing

A specialized document management application designed for handling preprinted forms. Forms processing systems often use multiple OCR engines and elaborate data validation routines to extract hand-written or poor quality print from forms to go into a database. With this type of application, it is essential to have good quality assurance mechanisms in place, since many of the forms that are commonly scanned were never designed for imaging or OCR.

Full-Text Indexing and Search

Enables the retrieval of documents by either word or phrase content. Every word in the document is indexed into a master word list with pointers to the documents and pages where each occurrence of the word appears.

Fuzzy Logic

A full-text search procedure that looks for exact matches as well as similarities to the search criteria, in order to compensate for spelling or OCR errors.

GIF

Graphics Interchange Format. CompuServe¨ 's native file format for storing images.

Gigabyte (GB)

2^{30} (approximately one billion) bytes, or 1024 megabytes. In terms of image-storage capacity, one gigabyte equals approximately 17,000 $8^{1}/_{2}$-inch x 12-inch pages scanned at 300 dpi, stored as TIFF Group IV images.

Grayscale

An option to display a black-and-white image file in an enhanced mode, making it easier to view. A grayscale display uses gray shading to fill in gaps or jumps (known as aliasing) that occur when displaying an image file on a computer screen.

Image Enabling

Allows for fast, straightforward manipulation of an imaging application through third-party software. For example, image enabling allows for launching the imaging client interface, displaying search results in the client and bringing up the scan dialogue box, all from within a third-party application.

Image Processing Card (IPC)

A board mounted in the computer, scanner or printer that facilitates the acquisition and display of images. The primary function of most IPCs is the rapid compression and decompression of image files.

Index Fields

Database fields used to categorize and organize documents. Often user-defined, these fields can be used for searches.

Internet Publishing

Specialized document management software that allows large volumes of paper documents to be published on the Internet or intranet. These files can be made available to other departments, offsite colleagues or the public for searching, viewing and printing.

ISIS and TWAIN Scanner Drivers

Specialized applications used for communication between scanners and computers.

ISO 9660 CD Format

The International Standards Organization format for creating CD-ROMs that can be read worldwide.

JPEG

Joint Photographic Experts Group (JPEG or JPG). An image-compression format used for storing color photographs and images.

Key Field

Database fields used for document searches and retrieval. Synonymous with index field.

MFP

Multifunction Printer or Multifunctional Peripheral. A device that performs any combination of scanning, printing, faxing, or copying.

Multipage TIFF

See TIFF.

Near-Line

Documents stored on optical discs or compact discs that are housed in the jukebox or CD changer and can be retrieved without human intervention.

OCR

Optical Character Recognition (OCR). A software process that recognizes printed text as alphanumeric characters. OCR enables full-text searches of documents and records.

Off-Line

Archival documents stored on optical discs or compact discs that are not connected or installed in the computer, but instead require human intervention to be accessed.

Online

Documents stored on the hard drive or magnetic disk of a computer that are available immediately.

Open Architecture

Applied to hardware or software whose design allows for a system to be easily integrated with third-party devices and applications.

Optical Discs

Computer media similar to a compact disc that cannot be rewritten. An optical drive uses a laser to read the stored data.

Pixel

Picture Element. A single dot in an image. It can be black and white, grayscale or color.

Portable Volumes

A feature that facilitates the transfer of large volumes of documents without the need to copy multiple files. Portable volumes enable individual CDs to be easily regrouped, detached and reattached to different databases for a broader information exchange.

Raster/Rasterized (Raster or Bitmap Drawing)

A method of representing an image with a grid (or map) of dots or pixels. Typical raster file formats are GIF, JPEG, TIFF, PCX, BMP, etc.

Record

Information, regardless of medium, that constitutes evidence of an organization's business transactions.

Record Series

A record series is a group of records subject to the same set of life-cycle instructions.

Region (of an image)

An area of an image file that is selected for specialized processing. Also called a zone.

Retention Period

The length of time that a record must be kept before it can be destroyed. Records not authorized for destruction are designated for permanent retention.

Scale-to-Gray

See Grayscale.

Scalability

The capacity of a system to scale up, or expand, in terms of document capacity or number of users without requiring major reconfiguration or re-entry of data. For a document management system to be scalable, it must be easy to configure multiple servers or add storage.

Scanner

An input device commonly used to convert paper documents into computer images. Scanner devices are also available to scan microfilm and microfiche.

SCSI Scanner Interface

The device used to connect a scanner with a computer.

Single-Page TIFF

See TIFF.

SQL

Structured Query Language. The popular standard for running database searches (queries) and reports.

Templates, Document

Sets of index fields for documents.

Thumbnails

Small versions of an image used for quick overviews that give a general idea of what an image looks like.

TIFF

Tagged Image File Format. A non-proprietary raster image format, in wide use since 1981, which allows for several different types of compression. TIFFs may be either single or multipage files. A single-page TIFF is a single image of one page of a document. A multipage TIFF is a large, single file consisting of multiple document pages. Document management systems that store documents as single-page TIFFs offer significant benefits in network performance over multipage TIFF systems.

TIFF Group III (compression)

A one-dimensional compression format for storing black and white images that is utilized by most fax machines.

TIFF Group IV (compression)

A two-dimensional compression format for storing black-and-white images. Typically compresses at a 20-to-1 ratio for standard business documents.

Versioning

In document or records management applications, the ability to track new versions of documents after changes have been made.

Workflow, Ad Hoc

A simple manual process by which documents can be moved around a multi-user document management system on an as-needed basis.

Workflow, Rules-Based

A programmed series of automated steps that routes documents to various users on a multi-user document management system.

WORM Disks

Write-Once-Read-Many Disks. A popular archival storage medium during the 1980s. Acknowledged as the first optical discs, they are primarily used to store archives of data that cannot be altered. WORM disks are created by standalone PCs and cannot be used on the network, unlike CD-Rs. In some industries, such as financial services, the definition of WORM has broadened to include other media, such as CD-ROMs and DVDs, which provide accessible but unalterable document storage.