Document Management for Law Offices

Mohammed Kashef
New York Law School

Institute of Information Law and Policy - Law Practice Technology Program

Introduction

- Law Offices have traditionally relied on paper filing systems for document storage and retrieval.
- However, paper records are extremely difficult to manage because they have to be stored in and retrieved from only one place.
- Electronic document management systems solve many of the storage and retrieval problems inherent in paper filing systems while simultaneously reducing business costs.
- EDMS manage storage and retrieval of many different types of digital documents, including word processing files, spreadsheets, database files, e-mail, voice mail, scanned images, and Internet/intranet HTML documents

What is a Electronic Document Management System

- Document Management Systems are technological solutions for storing, retrieving, and managing paper documents electronically
- Document Management Systems help law firms, law offices, and companies move into a relatively paperless environment where documents can electronically flow around the office without the need for hard copies.

Problem Presented

- Historically law firms and offices stored paperwork in rows upon rows of filing cabinets.
  - This management system required a lot of space and employees to handle the filing, retrieving, transferring of these documents.
  - Over periods of time
    - Files accumulates quickly and grow larger
    - Paper files are often hard to find.
    - Records may not be stored in their proper folder
  - Studies show that professionals often lose up to 500 hours a year looking for documents.

Problem Solved

- Electronic Document Management Systems offer a better way to manage records by enabling law offices\firms to manage millions of records and retrieve the one needed in seconds.
- EDMS:
  - User friendly
  - Allow documents to be shared:
    - Email
    - Fax
    - print
  - Protects confidential records
  - Prevents lost records
  - Saves storage space
  - Manages records easily
  - Finds documents quickly
  - Makes images centrally available
  - Eliminates the need for file cabinets
  - Enables disaster recovery

Components of a Document Management System

- Document Imaging System
- Document Workflow System

What is an Image

An Image is a visual representation of some measurable property of a person, object, or phenomenon.

What are Digital Images

- **DIGITAL IMAGES** are electronic snapshots taken of a scene or scan of a documents.
  - IE: photographs, manuscripts, printed texts, and artwork.
- Digital images are sampled and mapped as a grid of pixels.
- Each pixel is assigned a tonal value (black, white, shades of gray or color), which is represented in binary code (zeros and ones).
- The binary digits ("bits") for each pixel are stored in a sequence by a computer and often reduced to a mathematical representation (compressed). The bits are then interpreted and read by the computer to produce an analog version for display or printing.

The Document Imaging Process

- **Scanning**
  - Documents must first be scanned.
- **Indexing**
  - Then each document is marked with a specific category describing the type of document it is.
  - The documents are then arranged with other related documents that are grouped together.
- **Storage**
  - After the documents have been indexed they are kept for a specified period of time on hard disk.
  - Then they are later transferred into a more permanent storage place usually optical media.

Imaging of Electronic Documents

- Aside from scanned documents, an imaging system should be capable of capturing documents produced by the computer as text/ASCII files.
- Examples of computer generated ASCII files are purchase orders, work orders, checks, etc.

Scanning

- Imaging scanners should have Automatic Document Feeders.
- The proper scanner chosen must take into consideration the speed of the scanning
  - To decide on the proper scanning speeds budget, size, and volume of paper to be scanned should be evaluated.

Scanning Factors

- **Resolution/threshold** - Increasing resolution enables the capture of finer detail.
- **Bit Depth** - Increasing the bit depth, or number of bits used to represent each pixel, enables the capture of more gray shades or color tones.
- **Color** - Capturing and conveying color appearance is arguably the most difficult aspect of digital imaging.
- **System Performance -** The equipment used and its performance over time will affect image quality.

How Scanners Works

- Scanners operate by shining light at the object or document being digitized and directing the reflected light (usually through a series of mirrors and lenses) onto a photosensitive element.
- In most scanners, the sensing medium is an electronic, light-sensing integrated circuit known as a charged coupled device (CCD).
- The CCD convert levels of brightness into electronic signals that are then processed into a digital image.
- CCD is by far the most common light-sensing technology used in modern scanners. Two other technologies, CIS (Contact Image Sensor), and PMT (photomultiplier tube) are found in the low and high ends of the scanner market, respectively.

Types of Scanners

- Flatbed
- **Sheet-feed Scanners**
- **Drum Scanners**
- **Microfilm Scanners**
- **Slide Scanners**

Flatbeds

- Flatbed scanners are the best-known and largest selling scanner type.
- They're versatile, easy to operate, and widely available.
- Use a basic technology, in which a light sensor (generally a CCD) and a light source, both mounted on a moving arm, sweep past the stationary document on a glass platen.

Sheet-feed Scanners

- Sheetfeed scanners use the same basic technology as flatbeds, but maximize throughput, usually at the expense of quality.
- Generally designed for high-volume business environments, they typically scan in black and white or gray scale at relatively low resolutions.

Drum Scanners

- Drum scanners produce the highest resolution, highest quality scans of any scanner type, but at a price.
- Besides their expense, drum scanners are slow, not suitable for brittle documents and require a high level of operator skill.
- Typically found in service bureaus that cater to the color pre-press market.

Microfilm Scanners

- Microfilm scanners are highly specialized devices for digitizing roll film, fiche, and aperture cards.
- Getting good, consistent quality from a microfilm scanner can be difficult because they can be operationally complex, film quality and condition may vary, and because they offer minimal enhancement capability.
- Only a few companies make microfilm scanners, and the lack of competition contributes to the high cost of these devices.

Slide Scanners

Slide scanners are used to digitize existing slide libraries as well as photo intermediates of 3-dimensional objects and documents that are not well-suited for direct scanning.

Image Processing

- **Editing, touch-up, enhancement** – This includes steps such as descreening, despeckling, deskewing, sharpening, use of custom filters, and bit-depth adjustment.
- **Compression** —sometimes carried out by dedicated scanner firmware or dedicated hardware in the computer. Compression can also be a software-only operation though dedicated hardware is faster and should be considered when creating very large files or very large numbers of files.
- **File format conversion** — the original scan may not be in a format suitable for all intended uses, thus requiring conversion.
- **Scaling** — it's likely that scans captured at high resolution will not be suitable for on-screen display.
- **OCR** (optical character recognition)—conversion of scanned text to machine-readable text that can be searched or indexed.
- **Metadata creation** —addition of text that helps describe, track, organize, or maintain an image.

Indexing

- When paper documents are received for imaging, they must be organized so they can be easily retrieved.
- Paper Documents are labeled, sorted, stapled, placed in folders and filed in a cabinet. Without these steps, nothing could be found in a busy workplace.
- Electronic documents are no different. A document imaging system must have a comprehensive indexing system that organizes documents for future use.

Organizational Benefits of Indexing

- Indexing digital documents produces both tangible and intangible benefits to the organization.
- Tangible benefits include financial, legal, employee, and value-added benefits.
- Intangible benefits include less concrete measures of success, such as improved perception of the organization by both employees and customers.
- Combined tangible and intangible benefits result in financial gain for the organization through increased employee productivity, customer service, and competitive advantage in the marketplace.

Financial Benefits

- **Increased production** - The speed of many routine office procedures (such as production of statistical reports, records management tasks, access to and retrieval of digital documents, etc.) is increased.
- **Decreased future staff requirements** - Increases in production can be handled by current staff.
- **Increased access to current information** - Quick and accurate updates of indexes throughout the organization decreases information retrieval time and increases accuracy of information.
- **Improved customer service** - Prompt, accurate information retrieval increases repeat revenue for the organization.
- **Decreases in human filing mistakes** - Large legal practices often spend 8 or more hours to locate misfiled documents.

Methods of Indexing

- There are three different ways to index (organize) electronic documents in an imaging system:
  - Indexing works inside the document
  - Storing documents in folders
  - Assigning index fields to a document

Indexing Words Inside the Document

- Traditionally, keyword indexing has been used to make the information within a document available. Assigning key words from the document itself allows users to store and find pages later. Unfortunately, it can take a lot of time for qualified people to read and manually key word documents.

- Document imaging systems can eliminate the need for manual key word indexing by providing automatic full-text indexing. To do this, the software must have the capability to perform Optical Character Recognition (OCR). This process actually reads a scanned page and converts it into readable text. Once read, the imaging software can then automatically index every word to track the location of each word and phrase within every document, dramatically reducing indexing costs while providing improved searching capabilities.

Storing Documents in Folders

- Along with keyword or full-text indexing, an imaging system must have a visual method of filing documents.
- In any office, files are normally found by looking in a particular folder in a particular drawer in a particular file cabinet. An imaging system must have the ability to duplicate this filing system.
- A flexible folder structure eases the transition from paper filing to electronic filing and makes imaging systems more successful.

Assigning Index Fields to a Document

- The final method of organizing documents is through index fields or templates. An imaging system must use a robust index field structure to accommodate large volumes of documents. Generally, these structures are based on a database that maintains these index fields.
- Because of the need to integrate imaging systems with other applications, these databases must use industry-standard languages and tools such as SQL-compliant databases.

Storage

- Electronic images need a place to reside, and for the purposes of imaging, this place needs to be long-term, expandable and reliable.
- There are many different storage media available for imaging. Each one has its own strengths and weaknesses.
- For an imaging system, a good storage system must encompass changing technologies, increasing numbers of document volumes and the tests of time. Selecting the right medium depends on needs and budget restrictions.
- There are five storage options: (1)Magnetic Media, (2)Magneto-Optical Storage, (3)Compact Disks (4)DVDs, (5)WORM

Compression

- **COMPRESSION** is used to reduce image file size for storage, processing, and transmission. The file size for digital images can be quite large, taxing the computing and networking capabilities of many systems.
- All compression techniques abbreviate the string of binary code in an uncompressed image to a form of mathematical shorthand, based on complex algorithms.

TIFF Files

- TIFF is an acronym for Tag(ged) Image File Format. It is one of the most popular and flexible of the current public domain raster (A scanning pattern of parallel lines that form the display of an image projected on a cathode-ray tube of a television set or display screen.) file formats.
- TIFF is primarily designed for raster data interchange. It's main strengths are a highly flexible and platform-independent format which is supported by numerous image processing applications.

Magnetic Media

- With the fast response times and dramatic drops in hard drive prices, magnetic media such as hard drives or RAID (Redundant Array of Inexpensive Disks) have been an attractive solution for storage of document images.
- These devices are relatively inexpensive and can be linked together to store large numbers of documents. In addition, magnetic media provide the fastest response time.
- The problem with magnetic media is that while inexpensive, they still cost more than optical media and their moving parts are subject to mechanical failure. (but data is usually backed up)

Magneto-Optical Storage

- With the drops in hard drive prices, the attractions of magneto-optical storage are quickly fading. Magnetic Optical (MO) disks are reliable and can store large amounts of data.
- In addition, MO disks can be placed in a jukebox that can hold over a hundred disks at a time. MO technology is slower and more expensive than large hard drives.
- The drawbacks of MO technology include the expense and fragile nature of the media: As with magnetic drives, the information is written on a spinning platter, which can be erased or damaged. This type of damage would require restoration from backup.

Magneto-Optical Jukeboxes

HP

Plasmon

Compact Disks

- CDs offer a safe and reliable media that can provide long-term storage for images, in some cases up to 100 years.
- Disks can also be stored in jukeboxes that can hold 500 CDs at a time. Furthermore, CDs do not require any specialized hardware or software to retrieve information.
- The drawback with CDs is their limited storage capacity: A standard CD can only hold around 12,000 pages of documents (~1GB).

DVD

- Visually similar to CDs, these disks offer the same storage capacity of a MO disk without using moving parts in the media or requiring special software for decoding.
- With the life expectancy of CDs, DVD represents the best long-term option for reliable document imaging storage.

WORM

- The final storage medium is WORM (Write Once Read Many). This media format is not readily available and requires specialized hardware and software to operate.
- Because of the limited number of companies that provide materials and support for WORM technology, it is not highly recommended.

What is Workflow

Workflow is the automation of a business process, which documents are passed from one participant to another for actions, according to a set of procedural rules.

What is a Workflow System

- A workflow management system defines, creates and manages the execution of workflows, through use of software, running on one or more workflow engines, which is able to interact with workflow participants,and, where required, invoke the use of information technology (IT) tools and applications.

- Workflow management systems determine the flow of work according to predefined process definitions. It manages resources (i.e. applications, data, people) required to meet goals and provides monitoring facilities and control capabilities. The elimination of paper based procedures such as copying, manual archiving, and retrieval as well as in house distribution is often the most important economic argument for the introduction of workflow management system. Information relating to business operations is available instantly, giving management the opportunity to respond quickly to events or problems.

The Workflow Process

- **Retrieval**
  - The retrieval system uses information about the documents, including index and text, to find images stored in the system. A good retrieval system will make finding the right documents fast and easy.
- **Access**
  - Document viewing should be readily available to those who need it, with the flexibility to control access to system. A good access system will make documents viewable to authorized personnel, whether in the office, at different locations, or over the Internet.

Retrieval

- Once documents have been entered and indexed within an imaging system, rapid retrieval is a must. Users need to be able to use common sense tools to find any document within the system based on the most logical method. In some cases, this means using text, in other cases it would be based on the document folder or index field information. Whatever the method, document retrieval must be simple and user-friendly.
- Retrieval is where a powerful indexing system pays off. Users who are familiar with a document's text should be able to use that information to find what they want. A document imaging system must be able to use full-text retrieval because a person may not be familiar with the required key words
- Similarly, using the document name and folder view to find a document is also nice, but not always the best method. Once an imaging system contains thousands or millions of pages, folder trees become more complicated and document names become less unique. To assist searches, an imaging system needs to combine different criteria into one comprehensive search.
- The same is true for index field information. A full-featured imaging system will have user-definable template fields. Index field searches will allow a user to comb through millions of records in seconds to find the document necessary. Having the flexibility to combine template searches along with text and document names offers users the greatest control of their documents. A good imaging system makes retrieval of relevant documents fast, easy and efficient.

Access

- A full-featured document management system must provide the ability to permit access to those users who need it, without compromising security.
- To create this access, a system must have two fundamental features:
  - Broad availability
  - Comprehensive security

Broad availability

- An imaging system must offer different ways of accessing images. The most common method is through the user's desktop. Every document imaging system must provide a client-based user interface that enables the scanning, indexing and retrieval of documents. Without this basic interface, the system cannot function.

- To provide broad availability and access flexibility, document management systems must meet the requirements of offices with diverse uses and remote locations. Document management is no longer an "in-the-office" process. Many users require portability to exchange imaging information with other colleagues or to work off-site. An imaging system that does not offer this flexibility limits not only the usefulness of the system but also the abilities of the user.

Comprehensive security

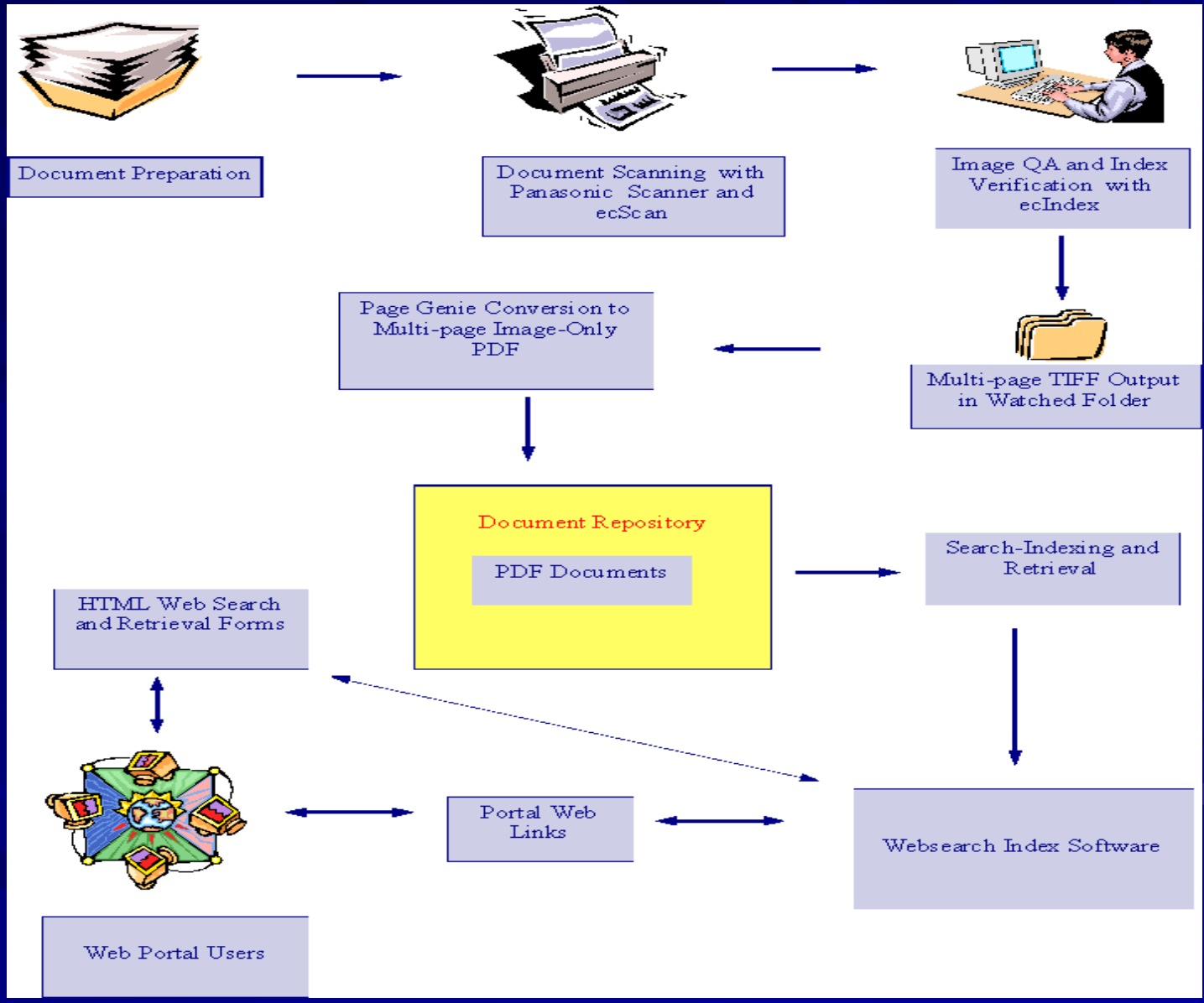
- The ability to provide imaging to a larger group means stronger control must be placed on user access. A comprehensive security system must allow the system administrator to control what users can or cannot do as well as what they can or cannot see. The system must control access to folders, documents in a simple and complete manner.
- A good access system will make document imaging available to everyone, whether they are in an office or at a remote location, all without compromising system security.

Web Access

- Document Management systems should be Web (HTML) based front end systems for managing and interacting with the information contained in electronic documents.
- When accessing the web based electronic documents they are viewed as PDF files. PDF portals as a essentially completely customized application interfaces for that meet document management and workflow requirements.

PDF Files

- **Portable Document Format** (**PDF**) is a file format developed by Adobe Systems for representing two dimensional documents in a device independent and resolution independent format.
- Each PDF file encapsulates a complete description of a 2D document that includes the text, fonts, images, and 2D vector graphics that compose the document.
- PDF is primarily the combination of three technologies:
  - a cut-down form of PostScript (a page description computer language) for generating the layout and graphics,
  - a font-embedding/replacement system to allow fonts to travel with the documents, and
  - a structured storage system to bundle these elements and any associated images into a single file, with data compression where appropriate.


Document Preparation
Document Scanning with Panasonic Scanner and ecScan
Image QA and Index Verification with ecIndex
Multi-page TIFF Output in Watched Folder
Page Genie Conversion to Multi-page Image-Only PDF
Document Repository
PDF Documents
Search-Indexing and Retrieval
HTML Web Search and Retrieval Forms
Portal Web Links
Websearch Index Software
Web Portal Users

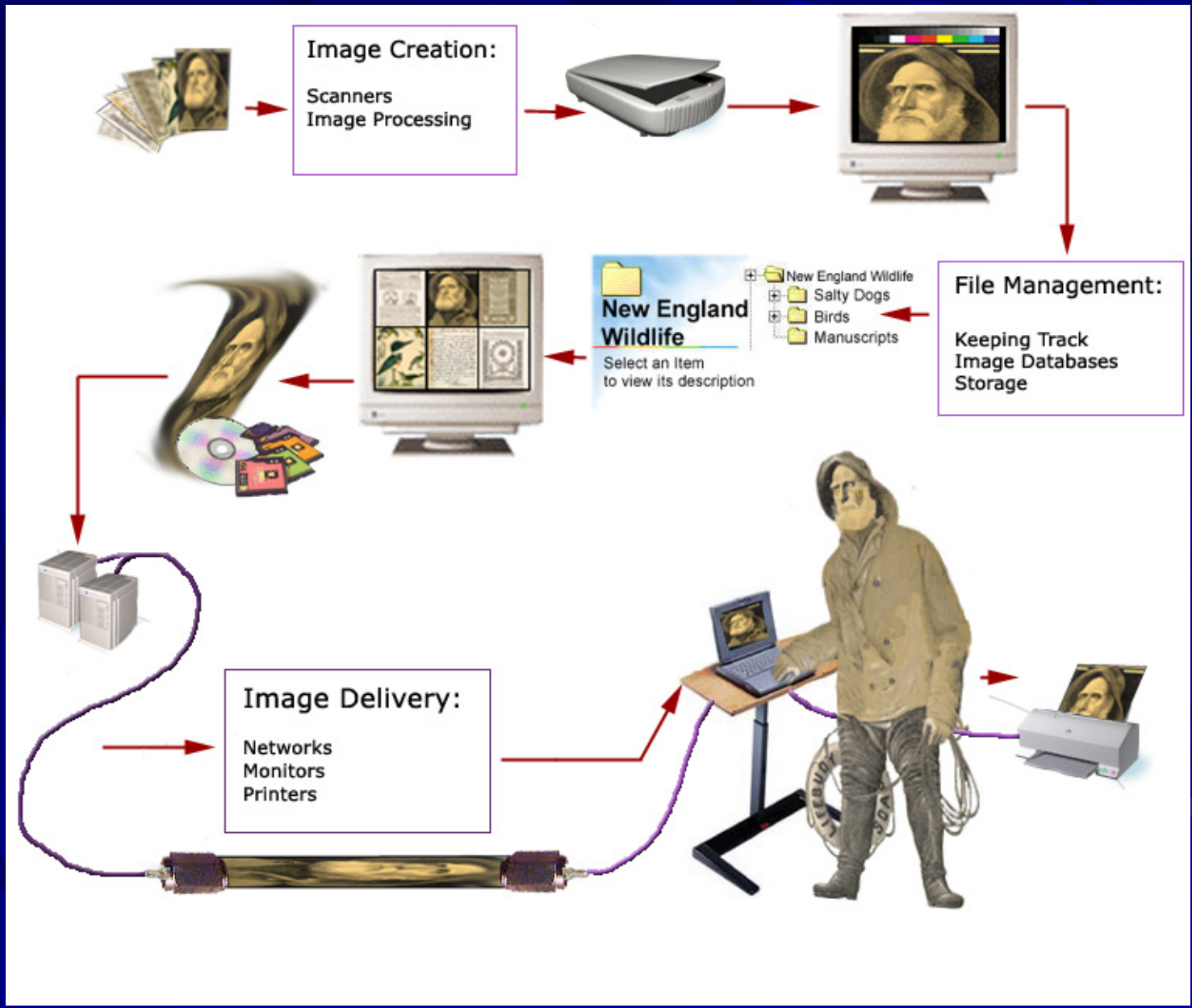


Image Creation:
Scanners
Image Processing
File Management:
Keeping Track
Image Databases
Storage
New England Wildlife
Select an Item
to view its description
New England Wildlife
Salty Dogs
Birds
Manuscripts
Image Delivery:
Networks
Monitors
Printers

Outsourcing vs. In-House

DETERMINING THE BEST APPROACH: OUTSOURCING VS. IN-HOUSE

- There are pros and cons to outsourcing or creating in-house capabilities for digital imaging efforts. Even when the decision is made to outsource certain functions, the institution must support many aspects of the process

Outsourcing

Outsourcing is viable if an institution has a good understanding of the near- and long-term goals of an imaging initiative, and can fully specify imaging, metadata, and derivative requirements; locate reliable vendors; evaluate products and services; adopt policies and procedures for various functions; and define institutional and vendor responsibilities.

Outsourcing Advantages

- Cost containment and limited risk;
- Costs typically lower than in-house figures, although prices vary widely
- Vendors can handle large volume and high throughput
- Expertise, training, technology obsolescence costs absorbed by vendor
- Broad range of options and services available, (including imaging, metadata creation, enhancements, processing, encoding, derivative creation, printing, storing and backup, database development)

Outsourcing Disadvantages

- Institution removed one step from imaging functions; services most often performed offsite or even off shore
- Vulnerability due to vendor instability
- Vendor inexperience with needs of cultural institutions
- Lack of standards and best practices with which to define requirements or negotiate for services
- Challenges in communication
- Security, handling, transportation issues

In-House Facility

Establishing an in-house facility requires an institution to support the full digitization chain with appropriate staff, space and facilities, equipment and supplies, and to absorb time and expense associated with ramping up.

In-House - Staffing

- ***Staff*** will be needed for the following tasks: identification, selection, preparation, digitization, metadata creation, quality control, cataloging, data loading, systems support, and management.
- Depending on the institutional configuration and the extent of the imaging program, staff will also need to be hired to develop and maintain the image database and Web delivery system.

In-House Hardware Requirements

- The facility must also provide the requisite communications—phone/data lines, LAN connections, and UPS (uninterrupted power supply) protection. It must support appropriate environmental controls, including proper HVAC, air filtration, and controlled lights (overhead and ambient). Scanning equipment and lights can raise temperatures, especially in confined areas. Consider workflow in designing the room configuration.

Hardware and Software Requirements

- Hardware
  - • Scanning devices
  - • High resolution monitors
  - • Workstations
  - • Peripherals
  - • Servers and storage devices
  - • Printers
- Software to support the following
  - • Operating system, networking/server/graphics support, programming packages
  - • Scanning, image editing, viewing, color management, quality control
  - • Derivative creation
  - • File management, workflow management
  - • Indexing
  - • Database management system
- Other equipment and supplies
  - • Copy stands/cradles/lights/lenses
  - • Quality control equipment and supplies
  - • Routine office supplies
  - • Storage media, paper, ink cartridges
  - • Documentation, technical manuals, reference publications

Considerations\Selection Criteria for building a EDMS

- **Legal Restrictions**
- **Document Attributes**
- **Preservation Considerations**
- **Organization and Available Documentation**
- **Intended Uses**
- **Digital Collection Building**
- **Duplication of Effort**
- **Institutional Capabilities**
- **Finances**

Legal Restrictions

- Begin your selection process by considering legal restrictions.
  - Is the material restricted because of privacy, content, or donor concerns?
  - Is it copyright protected?
    - If so, do you have the right to create and disseminate digital reproductions?

Document Attributes

- Does the material lend itself to digitization?
- Can the informational content be adequately captured in digital form?
- Do the physical formats and condition of the material represent major impediments?
- Are intermediates, such as microfilm or slides, available and in good condition?
- How large and complex in terms of document variety is the collection?

Preservation Considerations

- Would the material be put at risk in the digitization process?
- Would digital surrogates reduce use of the originals, thereby offering them protection from handling? Is the digital reproduction seen as a means to replace the originals?

Organization and Available Documentation

- Is the material in a coherent, logically structured order?
- Is it paginated or is the arrangement suggested by some other means?
- Is it complete?
- Is there adequate descriptive, navigational, or structural information about the material, such as bibliographic records or a detailed finding aid?

Intended Uses

- What kinds, level, and frequency of use are envisioned?
- Is there a clear understanding of user requirements?
- Can digitization support these uses? Will access to the material be significantly enhanced by digitization?
- Can your institution support a range of uses, e.g., printing, browsing, detailed review?
- Are there issues around security or access that must be taken into account

Digital Collection Building

- Is there added incentive to digitize material based on the availability of complementary digital resources (including data and metadata?)
- Is there an opportunity for multi-institutional cooperation?

Duplication of Effort

- Has the material already been digitized by another trusted source?
  - If so, do the digital files possess sufficient quality, documentation, and functionality to serve your purposes?
  - What conditions govern access and use of those files?

Institutional Capabilities

- Does your institution have the requisite technical infrastructure to manage, deliver, and maintain digitized materials?
- Do your principal users have adequate computing and connectivity to make effective use of these materials?

Finances

- Can you determine the total cost of image acquisition?
- Is this cost justified based on real or perceived benefits accruing from digitization?
- Are there funds to support this effort?
- Is there institutional commitment to the on-going management and preservation of these files?

Costs of Indexing

- Costs reported by companies indexing their documents in-house range from $.12 to $.20 per page.
- Typical service bureau charges currently range from $.15 to $.30 per page for scanning and indexing.

Costs of Updating

- The index itself must be kept current and updated.
  - direct costs of an "index maintenance" project average $.29 per document
  - Index maintenance may cost more that the original cost of indexing documents. T
- ime and effort spent on initial index design may eliminate costly projects to correct or update after the system is in place.

Legal Benefits:

- Litigation protection. In a lawsuit, records need to be produced very quickly.
  - An indexing system that can identify and retrieve documents needed for litigation can pay for itself if a single multi-million dollar lawsuit is avoided.
- Rule 26, which requires parties involved in a federal lawsuit to identify and produce relevant records within 85 days of the beginning of the litigation
- Records retention compliance. Federal, state and local governments regulate record retention periods for organizations. There are over 10,000 federal recordkeeping laws alone

Employee Benefits:

- Currency of business information.
  - New documents can be added to the indexing system quickly, and if documents are indexed when they are created, all users can access them immediately. Employees can do their jobs better.
- Document version control.
  - Indexing digital documents makes it possible to control which version of a document users can access. Employees don't waste time working on outdated documents, or updating a version that's already been revised.
- Remote access.
  - An organization-wide standard indexing language allows authorized users to retrieve documents from anywhere in the world. Employees don't have to take their whole office with them when they travel.
- Simultaneous access.
  - Employees can share a document if it is indexed properly and retrieved from a computer network. The "file folder" is never missing from the file cabinet. Hard copy production and distribution are also eliminated.
- Decreased training time.
  - New employees become quickly and fully productive in the organization

Conclusion

- Digital imaging technology is one of the most significant litigation support tool introduced to the legal profession in years. Firms that take advantage of this technology solution have performance and productivity gains that enable efficient management by allowing these offices to access information instantly
- EDMS provides quick turn around for documents received and are ready online within an hour
- EMNS provides searchable access under varied search methods - Searching through documents electronically allows a firm to be more efficient.
- Assistance in Litigation:
  - Large litigation can contain thousands and thousands of doc
  - On appeals everything is already scanned from the previous case
- Offices that do not use imaging have rooms stacked floor-to-ceiling with bankers boxes and file cabinets full of archives dating back to when law firms first opened.
  - This creates a real space problem and offices essentially can not get rid of paper to space becomes more and more of a problem over time and as more and more papers are gathered the more difficult it is to find the documents. Here it is essentially an electronic file cabinet
- Clients can share information easily over the internet
- Systems are user friendly, users such as paralegals should be able to use the system immediately and training should be relatively brief