Document Scanning Essentials

www.archives.nysed.gov

Objectives

• Learn about current imaging practices
• How to prepare for an imaging project
• How to incorporate RM principles into your project
• Provide strategies and best practices

Agenda

• Digital imaging overview
• Preparing for your project
• Current conversion practices & requirements
• Working with a scanning vendor
• Managing and preserving what was scanned
• Additional words of advice & wrap up
What is digital imaging?

**IMAGING OVERVIEW**

What is Imaging?

- Converting paper to electronic representation
  - Requires a scanner, software, PC and labor

Imaged Documents

- No native intelligence
  - Need additional descriptors (i.e. index) to find records
  - With no indexing, similar to a file folder with no label

- Need proper storage to manage scanned images
  - e.g. Electronic Content Management System (ECMS)
OCR/ICR for Access

- Optical Character Recognition (OCR)
  - Convert images to electronic text
  - 99.9% accuracy = ~20 errors per page
- Intelligent Character Recognition (ICR)
  - Convert hand-printing to electronic text
- Full text or zonal OCR
  - Entire page (minutes) or field-based OCR (voucher)
- Not always necessary

Storage and Retrieval

- Need software application to manage images
- Storing on network server without ECM software:
  - Harder to locate desired record
  - Users may be able to alter or delete records
  - May not notice until backup tapes are overwritten
  - Authenticity of records can be questioned
  - Less flexibility in restricting access

Why an ECMS?

- Database driven image management application
- Manage digital content & facilitate access
  - Provides security, audit trail, management, retention, and control of scanned images
  - Can manage native and other electronic content (e.g. spreadsheet, word processing, email, etc.)
When to Consider Imaging?
• High volume of actively retrieved records
• Need quick access
• Greater flexibility in finding
• Shared access across departments/locations
• Records being damaged from handling

When Not to Consider
• Never or infrequently accessed records
• Cannot afford to maintain image system
• No technical infrastructure to support ECMS

Popular Record Series
• Finance - Payroll register, general ledger
• Engineering - Project files; large format plans/maps
• Personnel - Employee files; benefit records
• Administrators - Contracts/agreements, audits
• County Clerks - Land records
• Town Clerks - Minutes, resolutions, ordinances
• Sheriff/Police - Incident/arrest reports, inmate files
• Social Services - Case files
• Building Dept - Permits
What records are you considering?

Do you have a system to manage your scanned records now?

Cost Benefit Analysis

• Results vary depending on:
  – Are you considering using vendor or in-house
  – Do you have equipment in place?
  – Do you have trained & available staff?

• Current storage/retrieval costs vs scanning costs
  – Staff time to store, retrieve and maintain files
  – Physical storage space costs
  – Offsite storage costs

Outsourcing Cost Estimate

• Cost to scan
  – $100-$180+ per box for standard office documents
  – Includes basic document preparation, 1-2 index fields

• Does not include:
  – Retrieval system (i.e. ECMS)
  – Retrieval of records during conversion project
  – Destruction of paper

• May or may not include:
  – Transportation of records
Outsourcing Cost Estimate

• If you were to have a vendor scan:
  – 50 boxes = $5,000-$9,000+
  – 250 boxes = $25,000-$45,000

• Does not include:
  – ECMS software, additional server storage, annual maintenance cost of system, or staffing

Storage Space Needed

• 1 scanned page = 50 KB
• 1 scanned box = 125 MB
• 1 scanned file cabinet = 1 CD-R
• 2 file cabinets = 8 boxes = 1 GB
• 10 file cabinets = 5 GB = 1 DVD-R
• 2,000 file cabinets = 8,000 boxes = 1 terabyte

Microfilm Alternative

• Strengths
  – Save space and money
  – Protect Information
  – Ensure file integrity

• Weaknesses
  – Hard to produce quality film
  – Difficult to add new records
  – Users don’t like it
Not an ‘Either/Or’ Decision

- Hybrid microfilm/imaging solution
- Meet two goals at the same time
  - Microfilm copy for preservation
  - Images for quick access

Imaging and/or microfilming?

Are you considering one or both?

Before you begin...

PREPARING FOR A PROJECT
Overall Process Steps

1. Identify records for the project
2. Identify who will do the project
3. The scanning/conversion process
4. Quality assurance
5. Transfer to your retrieval system
6. Users accessing images
7. Managing the stored images

Where to Start?

• Inventory your records
• Destroy obsolete/duplicate records
• Note remaining records’ characteristics
  – Volume (# of boxes, file drawers & type)
  – Document size
  – Single or double sided pages
  – General condition

Identify Records

• Specific records series targeted
  – Itemize and quantify
• Backfile conversion needed?
  – If so, how far back are records actively retrieved?
• Active records or day forward records only?
  – If there is no immediate need for older records
  – If backlog is too costly to tackle as a whole
Who Will Do the Project?

• Perform internally if you:
  – Have trained staff & equipment
  – Have a small backfile project
  – Have an ongoing need (day forward)

• Outsource if you:
  – Have a large or one time backfile project
  – Lack available, trained staff
  – Do not have equipment
  – Do not have space

In-house Needs

• Equipment
• Software
  – Scanning software
  – Electronic Content Management System (ECMS)
• Space for equipment and document preparation
• Trained staff
• Support for equipment and software

Types of Scanners

Flatbed
Sheet fed
Types of Scanners

Planetary

Wide Format

Digital Copiers

- Multi-function devices
  - Copier, scanner, fax
  - Easy of use
  - Not same quality or flexibility of desktop scanners
  - Can generate unnecessary & unmanaged copies

Cost of Scanners

- Personal/Low range (not recommended)
  - $100; Up to 2-10 pages per minute (ppm)
- Workgroup/Mid-range
  - $500 - $1,200; 15 to 35 ppm
- Departmental/High-speed
  - $800 - $9,000; 40 to 75 ppm
- Production
  - $7,000 - $25,000+; 80 to 130 ppm (260 ippm)
Cost of Scanners

• Flatbed book-edge scanners
  – $1,000 - $2,000
• Planetary
  – $2,000 – $120,000
• Wide format scanners
  – $2,000 - $18,000

Other Needs

• Software
  – Scanning and image management (e.g. ECMS)
• Workspace
  – Document preparation
• Workstation
  – Scanning and indexing
• Server & storage space
  – Depends on volume scanned and # of users

More Information

• You may need more info before deciding whether to scan in house or use a vendor
  – Outsourcing discussed later in this workshop
In house or using a vendor?

Which are you considering and why?

Scanning process overview and requirements

THE CONVERSION PROCESS

Project Management

- Project has to be managed (whether or not outsourced)
  - Scope
  - Requirements
  - Quality
  - Accuracy
  - Timeliness

- Who will be your project manager?
Conversion Process Steps

1. Document preparation
2. Image capture
3. Indexing
4. Quality control
5. Transfer to ECMS

Document Preparation

- Performed prior to scanning, including:
  - Sorting
  - Pulling staples & paper clips
  - Flattening
  - Removing duplicates and other unnecessary docs
- Time consuming, manual labor
- Upfront time reduces overall cost of scanning

Image Capture

- "True" processing throughput
  - Scanner speed is only small part of process
  - Advertised/Rated capacity vs. actual throughput
    - Raw scanning speed is with ideal equipment
    - Time to transfer & write each file to storage device is not included
    - Does not include doc prep, indexing or QA/QC
Image File Characteristics

- Non proprietary

- Acceptable scanned image formats
  - TIFF (ITU-T Group IV, formerly CCITT)
  - PDF/A preferred over PDF for long term retention
    - ISO standard, contains everything needed to render the document; not dependent on vendor

Image File Characteristics

- Compression
  - Lossless only

- Resolution
  - 300 dpi for office documents
  - 300-600 DPI for large format plans and maps

Indexing Plan

- Retrieval methods
  - Index fields
  - Text Search
  - File name (only if no ECMS is in use)

- Used to identify & retrieve records from ECMS
  - 2-4 index fields per record series is typical
  - Limit # of index fields to keep costs down
Sample Index Plan

• Payroll Registers
  – Document Type: Payroll Registers
  – Payroll Date: 02/15/2011
  – Disposition Year*: 2067

• Minutes
  – Document Type: Board Minutes
  – Meeting Date: 06/05/2012
  – Disposition Date*: Permanent
  – Text Search: (Via optional OCR)

• Sheriff’s Inmate file
  – Document Type: Inmate file
  – Inmate Last Name: Smith
  – Inmate first name: Michael
  – Inmate #: 10-39104
  – Release Date: 02/15/2011
  – Disposition Year: 2027

• Purchase Orders
  – Document Type: Purchase Orders
  – Purchase Order #: 11-2300122
  – Document Date: 02/15/2011
  – Disposition Date: 2007

• Personnel file
  – Document Type: Inactive Personnel file
  – Employee last name: Jones
  – Employee first name: William
  – Employee #: 7547
  – Termination Date: 06/25/2012
  – Disposition Year: 2019
Tips for the Capture Process

- Spend time designing the capture process
  - Techniques to ease process & increase accuracy
  - Document preparation
  - Document separators & barcoding
  - Pulling index data from existing systems

Quality Control

- Check images
  - Readability, contrast, resolution, orientation
  - Page count (images scanned = images stored)

- Check index
  - Typos, correct index to image, other errors

- Have other person perform quality control

Any changes in how you thought about proceeding with your imaging project?

Indexing, formats, documentation preparation, quality control, etc.?
Working with a Vendor

1. Write a requirements specification
2. Issue an RFP
3. Evaluate vendors
4. Choose a vendor
5. Develop a contract
6. Monitor project actively

Requirements Specification

- Name of organization
- Name and arrangement of records series
- Total number of images
- Condition of records
- Preparation requirements
- Required image format
- Required indexing
- Other requirements (technical or process)
Issue RFP

• Incorporate the defined requirements
• Have vendors view the records
• Answer questions
• Issue clarification/RFP addendum if needed

Evaluating Vendors

• How many years of experience do you have?
• What types of projects have you done?
• Can you provide local government references?
• Does price quote match requirements?

Evaluating Vendors

• What are your facilities like?
  – Physical conditions: secure and protected
  – Staff: knowledgeable and careful with records
• What quality controls are in place?
  – Conduct visual and technical inspections
• What is the turn-around time?
• How do you handle errors?
Add to Standard Contract

- Delivery dates
- Packing, pickup, and delivery
- Contract resolution and cancellation
- No subcontracting clause
- Image transfer process and assistance

Work Does Not End Here

- Selecting a vendor does not mean work is done
  - Project still has to be managed
  - Periodic review points must occur
  - Receive sample images and index to test
  - Refine and adjust as needed

Do you have concerns using a vendor?

Any experiences you want to share?
Managing what was scanned

STORAGE & ACCESS

Storage Options

• Keep on server-class hardware (not desktop PC)
  – Network server-based stored
  – Network Attached Storage (NAS)
  – Storage Area Network (SAN)
• Do not keep on CD-R/DVD-R for active retrieval
• Backup images/index and store off-site
  – External hard drive
  – Use magnetic tape for backup purposes only

Access Options

1. ECMS
   – Index specific key data fields
   – Full-text searching
2. No ECMS
   – Manual access to e-file folders
3. Vendor provided standalone search tool
   – Search tools delivered with images
   – Limited to specific media delivered on
   – Usually single user license
Retention Issues

• Retention period
  – Determine ‘record copy’

• Media life expectancy

• Information systems stability

Image Transfer

• Getting the images to your system
  – Images
  – Index
  – Import script (optional)

• Transport via
  – Removable media (USB drive, CD/R, DVD/R)
  – External hard drive (fastest transfer)
  – Download from vendor’s site (small volumes)

Importing to ECMS

• Copy from transfer media to server
• Import into ECMS
• Review Import log for errors
• Plan for Q/A time
• Include vendor assistance in contract
Retaining/Preserving

• Commitment to maintain system
  – Annual maintenance costs
  – Support
  – Upgrades
• Must plan for future migration
  – At some point, current technology will be replaced
• Backup and off site storage
  – Business continuity must be considered

Paper Records

• Retain until Quality Assurance is 100% complete

• Similar to a microfilming project

Have your thoughts changed?

Has your approach to your project changed during this workshop?
LAST WORDS OF ADVICE

Wrap Up

- Set clear goals
  - What record series, volume, document characteristics
- Consider all solutions and overall costs
- Obtain needed resources
- Address records management issues
- Manage change

Document Scanning Essentials

Thank You!

For more information:
www.archives.nysed.gov