



CATALOGING
MAINTENANCE CENTER
Statewide Cataloging Support

So, You Want to Start a Digitization Project

Kat Anderberg, CMC Metadata Cataloger

Online With the CMC—December 11, 2025

Agenda

- Why We Digitize
- What to Digitize
- Copyright, Public Domain, and Other Legal and Ethical Considerations
- Digital Images 101
- Selecting Equipment
- Estimating Costs
- Creating Workflows
- Digitization Guidelines
- Storage
- Digital Preservation 101
- Resources and Questions

Why Digitize?

- Availability
- Accessibility
- Preservation



Teachers' strike in Dupu, IL, September 1979

What to Digitize?

- Intellectual or cultural value of the collection to researchers
- Demand from current or potential users
- Historical or geographical area covered by the collection
- Has another institution digitized the same (or similar) materials?
- Physical condition of the collection
- Is the material suitable for digitization?
- Will preservation work need to be done prior to digitization?
- Bound volumes and large materials
- Copyright permission

(Selection Criteria from ISL Digital Imaging Program Best Practices, Rev. 2018)

What to Digitize?

DIGITAL PROJECT PRIORITIZATION

	No	Maybe/ Unsure	Yes	Absolutely!
Value: Is there interest or value in these materials for genealogists, educators, researchers, community members, or other identified audiences?	1	2	3	4
Value: Do the materials contribute new voices or perspectives to the historical record (for example, materials document historically underrepresented groups in our community)?	1	2	3	4
Information context: Is there enough information available to add useful context (we know or can find out names of people, locations, dates)?	1	2	3	4
Availability: Are the materials unique and not already available online?	1	2	3	4
Legal/Ethical issues: Are the materials in the public domain (or we can get permission from the copyright holder) and there are no privacy concerns or other barriers to putting them online?	1	2	3	4
Condition: Are the materials in high-risk, deteriorating formats, particularly audiovisual recordings on media like audiocassettes, VHS tapes, CDs, or DVDs?	1	2	3	4

TOTAL
SCORE:



CATALOGING
MAINTENANCE CENTER
Statewide Cataloging Support

Copyright and Public Domain

- Published in the US before 1930: in the public domain
- Published in the US 1930-1977 **without** a copyright notice: in the public domain
- Published in the US 1978-1 March 1989 **without** a copyright notice and **without** subsequent registration within 5 years: in the public domain
- Published in the US 1930-1963 **with** copyright notice **but** copyright was **not** renewed: in the public domain

Copyright and Public Domain

- Published abroad before 1930: in the public domain
- Published abroad 1930-1977 **without** compliance with US formalities and are in the public domain in source country as of 1 January 1996: in the public domain
- Unpublished works from authors who died before 1955: in the public domain
- Unpublished anonymous works created before 1905: in the public domain
- Unpublished works created before 1905 when the date of the author's death is unknown: in the public domain

Copyright Permissions

- Works that are not in the public domain require copyright owner's permission to digitize
- Is the copyright owner findable?
 - If the copyright owner is deceased, is their next of kin findable?

Other Legal and Ethical Considerations

- Privacy issues
- PPI
- Cultural sensitivity
- Personally identifiable information
- 100-year rule

Digital Images 101

- Resolution
- Image Color
 - Bitonal
 - Grayscale
 - RGB
- Bit Depth
- File Formats
- Compression
- File Naming
- Color Management
- OCR

Resolution



Band members at Dupu Community High School, 1941-1942

8x10 inch print x 300 PPI = 2400x3000 pixels



Diplomatic gift presented to Admiral Carlisle A. H. Trost

2x2 inch print x 600 PPI = 1200x1200 pixels

Image Color and Bit Depth

- Bitonal
- Grayscale
- RGB
- CMYK*
- 1 bit (black and white)
- 8-bit grayscale
- 16-bit grayscale
- 8-bit color
- 24-bit color
- 48-bit color

File Formats and Compression

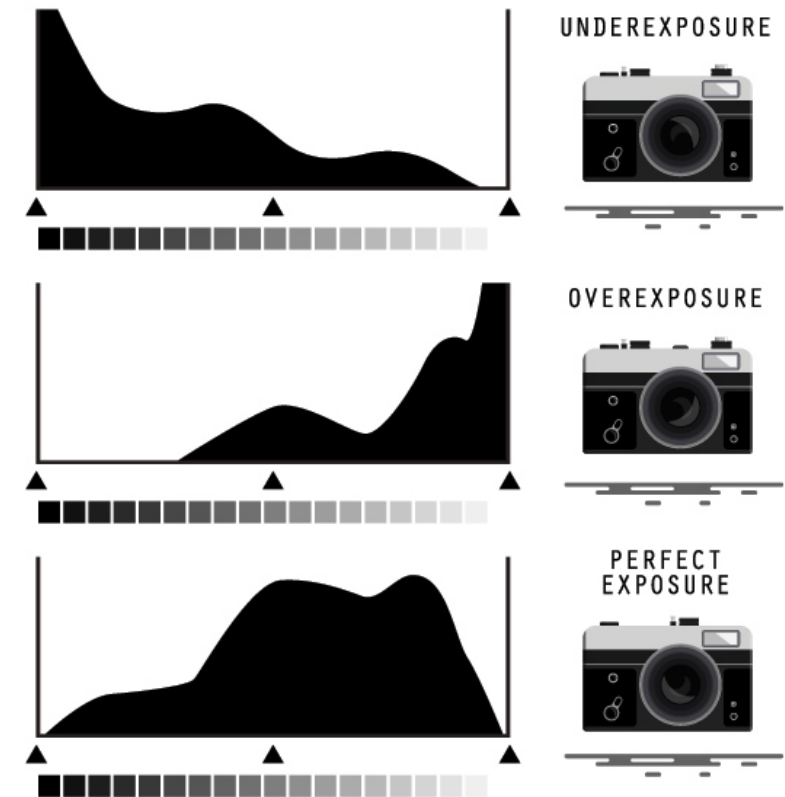
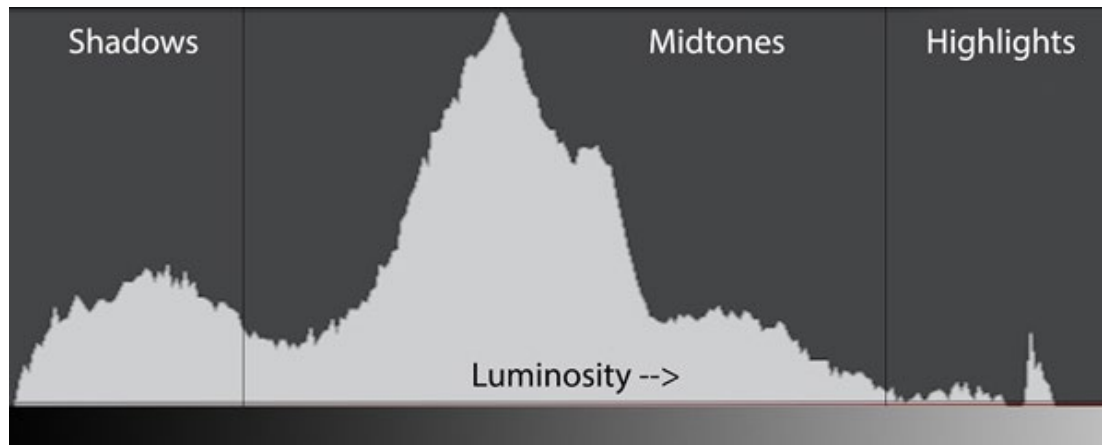
- Lossless vs. Lossy Compression
- Common File Formats:
 - TIFF
 - JPEG
 - PNG
 - RAW

File Naming

- Unique
- Consistent
- Technical constrictions
 - Leading zeros
 - Lowercase letters and numerals 0-9
 - Avoid punctuation EXCEPT for underscores and hyphens
 - Limit to 32 characters, including file extension (three characters)

Color Management

- Calibration
- Color Targets
- Histograms



Histogram illustrations from <https://photographypro.com/histogram>

OCR (Optical Character Recognition)

- Necessary for accessibility
- Images of Text \neq Actual Text
- Higher quality scan = Higher quality OCR
- You can't scan what isn't there
 - Faded text
 - Printing errors
 - Other damage

Equipment: Factors to Consider

- Material types
- Size of materials
- Quantity
- Condition
- Staff experience
- Number of staff
- Budget
- Physical space available
- Project duration
- End use of materials

Scanners—Flatbed vs. Sheet-Fed vs. Digital Camera

	Flatbed Scanner	Sheet-Fed Scanner	Digital Camera
Pros	<ul style="list-style-type: none">• Affordable• Can digitize many different types of materials• High resolution• Easy to use• Often come with templates for slides, negatives, etc.	<ul style="list-style-type: none">• Faster• Often the resolution is as good as flatbed scanners• Automation	<ul style="list-style-type: none">• Can digitize many different types of materials, including 3-D and oversized materials• Options for different lenses• Generally good image quality
Cons	<ul style="list-style-type: none">• Labor intensive• Slow• Cannot digitize 3-D materials• Difficulty with larger materials	<ul style="list-style-type: none">• Unsuitable for many types of materials (fragile materials, bound materials, 3-D materials)• More expensive than flatbed scanners• Difficulty with larger materials	<ul style="list-style-type: none">• Higher quality models are expensive• High learning curve

Picking a Scanner

- Image quality
- Reliability
- Performance
- Functionality
- Cost



Resolution Redux: Optical vs. Interpolated



DS640

Compact Mobile Document Scanner

- Ultra-compact and mobile
- Fast scan speeds of up to 16ppm† for both color and B&W
- Powerful software increases document efficiency

★★★★★ 4.1 (179) [Write a review](#) [Ask a question](#)

Original Price ~~\$134.99~~

Instant Rebate From Brother **Limit (2) per customer** - \$30.00

Your Price **\$104.99**

**Interpolated Scan
Resolution (max. dpi)**

Up to 1200 x 1200 dpi

**Optical Scan Resolution
(max. dpi)**

600 X 600



CATALOGING
MAINTENANCE CENTER
Statewide Cataloging Support

Who's Going to Do All This?

	DIY	Vendor
Pros	<ul style="list-style-type: none">• Direct control• You own the equipment!• You can scan material in batches	<ul style="list-style-type: none">• They own the equipment!• You can send materials in batches
Cons	<ul style="list-style-type: none">• Costs money (staff time)• You own the equipment! And are therefore responsible for maintaining it	<ul style="list-style-type: none">• Costs money (dollars)• Fragile items—do you trust the vendor to handle them appropriately?

How Much is This Going to Cost?

Number of scans ?

6000

Staffing

Librarian

Salaried

50000

0

— Remove Staff

Library Assistant


Hourly


20


0

+ Add Staff


How Much is This Going to Cost?


 Capture Device


Capture Device 


Flatbed scanner (i.e., an Epson 11000XL) 


Performed By


Librarian 


 Preparation of Original Materials

 Quality Control

 Post Processing

 Descriptive Metadata Creation

 Post Preparation

 Other

Previous

Skip

Save and continue

How Much is This Going to Cost?

Digitization Estimates

Number of scans: 6,000 scans
Total Estimated Time: NaN hours
Total Estimated Cost: \$0.00

Staffing

Librarian salaried \$50000 0%

Library Assistant hourly \$20 0%

Task	Sub-Task	% Materials	Performed By	Time (in minutes)	Cost
Image Capture	Flatbed scanner (i.e., an Epson 11000XL)		Librarian	11392.00	\$4,556.80
Total:				NaN	\$0.00

Digitization Estimates

Number of scans: 6,000 scans
Total Estimated Time: NaN hours
Total Estimated Cost: \$0.00

Staffing

Librarian salaried \$50000 0%

Library Assistant hourly \$20 0%

Task	Sub-Task	% Materials	Performed By	Time (in minutes)	Cost
Image Capture	Flatbed scanner (i.e., an Epson 11000XL)		Library Assistant	11392.00	\$3,759.36
Total:				NaN	\$0.00

How Much is This Going to Cost?

Digitization Estimates

Number of scans: 6,000 scans

Total Estimated Time: 6740.27 hours

Total Estimated Cost: \$122,061.98

Staffing

Librarian salaried \$50000 0%

Library Assistant hourly \$20 0%

Task	Sub-Task	% Materials	Performed By	Time (in minutes)	Cost
Image Capture	Flatbed scanner (i.e., an Epson 11000XL)		Library Assistant	11392.00	\$3,759.36
Preparation of Materials	Condition Review	100%	Librarian	729.75	\$291.90
Preparation of Materials	Fastener Removal	20%	Library Assistant	524.36	\$173.04
Preparation of Materials	Rights Review	100%	Librarian	233.40	\$93.36
Preparation of Materials	Sort Materials Into Items	50%	Library Assistant	933.24	\$307.97

Post Processing	Alignment	20%	Library Assistant	354.72	\$117.06
Post Processing	Clean Up	20%		0.00	\$0.00
Post Processing	Color Correction	20%	Library Assistant	1152.13	\$380.20
Post Processing	Cropping	20%	Library Assistant	714.20	\$235.69
Post Preparation	Desorting	0%		0.00	\$0.00
Quality Control	Level 1	100%	Librarian	1887.72	\$755.09
Descriptive Metadata	Level 1	75%	Librarian	386494.38	\$115,948.31
Total:				404415.91	\$122,061.98

Go With the (Work)Flow

- Selecting materials
- Evaluation their condition
- Cataloging
- Creating metadata
- Scheduling production
- Digitization prep
- Digitization
- Post processing
- Quality control review
- Archiving
- Publishing

Go With the (Work)Flow

- Key elements of a digitization workflow
 - Create a primary file (archival file)
 - Name the file according to your naming scheme
 - Conduct quality control review
 - Create an access copy/user file
 - Move primary file to stable storage
- Questions to ask yourself
 - Who is responsible for what?
 - Where are the files stored?
 - What challenges might you face, and how will you solve them?

Quality Control

- Size
- Resolution
- Color mode
- Bit depth
- Orientation
- Pixelation
- Focus
- Dust
- Details
- Cropping

Stay In the (Guide)Lines—FADGI

3.8 Prints and Photographs

Performance Level:	1-Star	2-Star	3-Star	4-Star
Master File Format	TIFF, JPEG 2000	TIFF, JPEG 2000	TIFF, JPEG 2000	TIFF, JPEG 2000
Access File Formats	All	All	All	All
Resolution (Sampling Frequency) (Units are Pixels Per Inch/ppi minus Reproduction Scale Accuracy)	≥ 194ppi (200 ppi – 3%)	≥ 243.75ppi (250ppi – 2.5%)	≥ 392ppi (400 ppi – 2%)	≥ 594 ppi ⁴ (600 ppi – 1%)
Bit Depth	8	8	8 or 16	16
Color Space	Gray Gamma 2.2, sRGB, Adobe RGB (1998), ProPhoto, ECIRGB_v2	Gray Gamma 2.2, sRGB, Adobe RGB (1998), ProPhoto, ECIRGB_v2	Adobe RGB (1998), ProPhoto, ECIRGB_v2	Adobe RGB (1998), ProPhoto, ECIRGB_v2
Color Mode	Grayscale or Color	Grayscale or Color	Color	Color

Stay In the (Guide)Lines—IDA

ILLINOIS DIGITAL ARCHIVES GENERAL GUIDELINES FOR IMAGES			
MEDIA	SCAN FORMAT	MASTER IMAGE	ACCESS IMAGE
Printed text	8-bit grayscale	TIFF Uncompressed Resolution: 300 ppi	JPEG Medium quality compression Resolution: 150 ppi (grayscale) Resize images to 1024 pixels across long dimension
Handwritten text	8-bit grayscale or 24-bit color	TIFF Uncompressed Resolution: 300 ppi	JPEG Medium quality compression Resolution: 150 ppi Resize images to 1024 pixels across long dimension
B/W photo	8-bit grayscale or 24-bit color	TIFF Uncompressed Resolution: 300 ppi	JPEG Medium quality compression Resolution: 150 ppi Resize images to 1024 pixels across long dimension
Color photo Color slide B/W slide Map Illustration etc.	24-bit color	TIFF Uncompressed Resolution: 300 ppi	JPEG Medium quality compression Resolution: 150 ppi Resize images to 1024 pixels across long dimension
3-D artifact	24-bit color	TIFF Uncompressed Resolution: 300 ppi	JPEG Medium quality compression Resolution: 150 ppi Resize images to 1024 pixels across long dimension
Newspapers	8-bit grayscale or 24-bit color	TIFF Uncompressed Resolution: 300 ppi	JPEG Medium quality compression Resolution: 150 ppi As a general rule, do not resize the image

Stay In the (Guide)Lines—Digital Readiness Toolkit

Content Type	Resolution (PPI)	Color	Bit Depth	File Format/Extension	Details
Books or other texts with no images	Min: 300 Ideal: 600	B & W or Grayscale	1 (bitonal) or 8	Min: PDF/A Ideal: TIFF	Optional Character Resolution (OCR) applications work best on documents scanned at 400 DPI or higher. Access files: PDF/A.
Books or other texts with images	Min: 300 Ideal: 400	Min: Grayscale Ideal: Color	Min: 8 Ideal: 24	TIFF	Optional Character Resolution (OCR) applications work best on documents scanned at 400 DPI or higher. Access files: PDF/A.
Manuscripts, letters, diaries	Min: 300 Ideal: 600	Color	Min: 8 Ideal: 24	TIFF	Aim for 3000-4000 pixels on the longest edge of the digital file.
Slides, film and negatives smaller than 4"x5"	Min: 1200 Ideal: 3000	Min: Grayscale Ideal: Color	Min: 16 Ideal: 24	TIFF	Aim for 3000-4000 pixels on the longest edge of the digital file.

Photos and prints smaller than 8"x10"	Min: 400 Ideal: 600	Min: Grayscale Ideal: Color	Min: 16 Ideal: 24	TIFF	Aim for 3000-4000 pixels on the longest edge of the digital file.
Photos, posters, and maps larger than 8"x10"	Min: 300 Ideal: 600	Min: Grayscale Ideal: Color	Min: 8 Ideal: 24	TIFF	Aim for 3000-4000 pixels on the longest edge of the digital file.
Audio	96 kHz		Sample rate: (recommended) 24-bit, 96kHz, (minimum) 24-bit, 48kHz	Broadcast wave (BWF), .wav or .aif	Access files: 320 Kbps sampling rate. File format: mp3 Codec: LPCM
Video (transferred from analog source)	720x486 4k		8-10 uncompressed Sample rate: (recommended) 24-bit, 48kHz, (minimum) 16-bit, 48kHz	.mov or .avi (Quicktime) MKV (Matroska)	Access files: 1.5 Mbps sampling rate. File format: mp4 h.264 Codec: Uncompressed 4:4:4:4 (motion picture film) Uncompressed 4:2:2 (video) Prores 4444 Codec (audio): PCM



Where Are All These Files Going to Live?

- Local Storage
 - Hard drives
 - Servers
 - Network Drives
- Storage Partners
 - Trusted organizations that work together by storing files for each other
 - There is usually a service agreement
- Vendors
 - Cloud Storage

How Much Storage Will You Need?

- 6,000 files x 50 MB x 3 copies = 900,000 MB (900 GB)
- 6,000 files x 50 MB x 2 copies = 600,000 MB (600 GB)

- 6,000 files x 15 MB x 3 copies = 270,000 MB (270 GB)
- 6,000 files x 15 MB x 2 copies = 180,000 MB (180 GB)

Digital Preservation 101

- Obsolete software
- Obsolete hardware
- Accidental deletion or saving over the original
- Viruses
- Physical and natural disasters
- File corruption
- Protect your investment!!

Digital Preservation 101

- Fixity checks
- Media migration
- Format migration

Digital Preservation 101

- 3-2-1 Rule
 - Have 3 copies of your files
 - Stored on at least 2 different types of media
 - At least 1 copy stored off-site
- Where are your files located?
 - Who has access to them?
- Inventory

Digital Preservation 101—Checksums

- Document
 - HashMyFiles
 - DROID
 - Data Accessioner
- Document AND Monitor
 - MD5 Summer
 - Exact File
 - Corz

Levels of Digital Preservation

Functional Area	Level			
	Level 1 (Know your content)	Level 2 (Protect your content)	Level 3 (Monitor your content)	Level 4 (Sustain your content)
Storage	<p>Have two complete copies in separate locations</p> <p>Document all storage media where content is stored</p> <p>Put content into stable storage</p>	<p>Have three complete copies with at least one copy in a separate geographic location</p> <p>Document storage and storage media indicating the resources and dependencies they require to function</p>	<p>Have at least one copy in a geographic location with a different disaster threat than the other copies</p> <p>Have at least one copy on a different storage media type</p> <p>Track the obsolescence of storage and media</p>	<p>Have at least three copies in geographic locations, each with a different disaster threat</p> <p>Maximize storage diversification to avoid single points of failure</p> <p>Have a plan and execute actions to address obsolescence of storage hardware, software, and media</p>
Integrity	<p>Verify integrity information if it has been provided with the content</p> <p>Generate integrity information if not provided with the content</p> <p>Virus check all content; isolate content for quarantine as needed</p>	<p>Verify integrity information when moving or copying content</p> <p>Use write-blockers when working with original media</p> <p>Back up integrity information and store copy in a separate location from the content</p>	<p>Verify integrity information of content at fixed intervals</p> <p>Document integrity information verification processes and outcomes</p> <p>Perform audit of integrity information on demand</p>	<p>Verify integrity information in response to specific events or activities</p> <p>Replace or repair corrupted content as necessary</p>
Control	<p>Determine the human and software agents that should be authorized to read, write, move, and delete content</p>	<p>Document the human and software agents authorized to read, write, move, and delete content and apply these</p>	<p>Maintain logs and identify the human and software agents that performed actions on content</p>	<p>Perform periodic review of actions/access logs</p>
Metadata	<p>Create inventory of content, also documenting current storage locations</p> <p>Backup inventory and store at least one copy separately from content</p>	<p>Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)</p>	<p>Determine what metadata standards to apply</p> <p>Find and fill gaps in your metadata to meet those standards</p>	<p>Record preservation actions associated with content and when those actions occur</p> <p>Implement metadata standards chosen</p>
Content	<p>Document file formats and other essential content characteristics including how and when these were identified</p>	<p>Verify file formats and other essential content characteristics</p> <p>Build relationships with content creators to encourage sustainable file choices</p>	<p>Monitor for obsolescence, and changes in technologies on which content is dependent</p>	<p>Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed</p>



Digital Readiness Levels

Focus Area	Bronze	Silver	Gold
	<i>Laying a strong foundation</i>	<i>Putting it into practice</i>	<i>Refining and sustaining</i>
Plan and Prioritize	Set goals for digital work that fit the organization's mission and policies. Adopt a digital mission statement or revise existing mission statement to include digital work.	Identify and prioritize potential digital projects. Make a digital project plan that includes roles, activities, required resources and partners.	Adopt a digital collection development policy or revise existing policy to include digitized and born digital content.
Obtain Permissions	Create and use permission forms and donor agreements that include specific language for the use of digitized and born digital content or modify existing forms.	Evaluate copyright status of content. Identify items with access restrictions or concerns, including privacy, ethical, or cultural considerations.	Assign standardized rights statements or Creative Commons licenses for collection items. Adopt a takedown policy and, if applicable, a statement on harmful content.
Digitize	Determine standards and procedures to be used to digitize physical materials or process born digital content.	Using identified standards, undertake digitization or born-digital processing work either in-house or with an appropriate vendor or partner.	Use a quality control checklist to review content and confirm it meets identified standards.
Describe	Adopt a consistent naming convention for digital files. Determine standards to be used to describe digital content.	Using identified standards, create basic descriptive metadata for terms.	Develop a data dictionary and use controlled vocabularies to standardize metadata.

Share	Review goals and options for providing access to content. Choose an access platform or system that meets identified needs.	Make items and associated descriptive information available for discovery and repurposing.	Implement techniques to support accessibility of online content, including alt text, transcripts, and other best practices.
Store and Maintain	Create and maintain a collection-level inventory of digital content.	Store at least two, preferably three, copies of each primary file and related metadata, with one copy stored off-site. Check and refresh storage media on a regular schedule.	Plan for future storage needs. Use software tools to check file integrity.
Evaluate	Identify primary users and ways to engage them with digital content.	Collect data and stories about how digital content is stored.	Use collected data and stories to inform future collection development, outreach, and programming. Share knowledge with other practitioners to build community around digital work.



CATALOGING
MAINTENANCE CENTER
Statewide Cataloging Support

Resources

- [Home - Digitizing Collections - LibGuides at Atla](#)
- [Home - Creating Digital Collections - LibGuides at University of Texas at Austin](#)
- [ISL Digital Imaging Program Best Practices, Rev. 2018](#)
- [Digital Project Planning Worksheet – Digital Readiness Toolkit](#)
- [Copyright Term and the Public Domain - Copyright Services - LibGuides at Cornell University](#)
- [RightsStatements.org](#)
- [Rights Statement Selection Tool](#) (PA Digital)
- [PII considerations in screening archival records](#) (archives.gov)
- [Bulk Rename Utility](#) (for Windows)

Resources

- [File Naming - Digital Projects Toolkit - LibGuides at North Dakota State Library](#)
- [Best Practices for Naming Electronic Records | Wisconsin Historical Society](#)
- [Guidelines: Technical Guidelines for Digitizing Cultural Heritage Materials - Federal Agencies Digital Guidelines Initiative](#)
- [Guidelines for Digitizing Archival Materials for Electronic Access | National Archives](#)
- [External Digitization Standards | Society of American Archivists](#)
- [Focus Area 3: Digitize – Digital Readiness Toolkit](#)
- [Recommended Formats Statement - Resources \(Preservation, Library of Congress\)](#)

Resources

- [Recommended file formats - eCommons: Cornell's Digital Repository - LibGuides at Cornell University](#)
- [Best Practices for the Selection of Electronic File Formats | Wisconsin Historical Society](#)
- [Histogram: Discover How To Take Better Photos By Exposing To The Right](#)
- [ISA Golden Thread Object – Level Target 1x | DT Heritage](#)
- [PDF7: Performing OCR on a scanned PDF document to provide actual text | Techniques for WCAG 2.0](#)
- [Digitization Dashboard](#)
- [Feet On The Ground: A Practical Approach To The Cloud Nine Things To Consider When Assessing Cloud Storage](#)

Resources

- [Plan Ahead for Disasters | Ready.gov](#)
- [Planning Guides | FEMA.gov](#)
- [Emergency Management \(Preservation, Library of Congress\)](#)
- [Disaster Response and Recovery | National Archives](#)
- [Focus Area 6: Store and Maintain – Digital Readiness Toolkit](#)
 - [Appendix C: Audiovisual Collection Inventory and Instructions – Digital Readiness Toolkit](#)
- [Digital Preservation Inventory Template for Museums - Canada.ca](#)
- [Digitization - Preservation - LibGuides at American Library Association](#)
- [What is Fixity, and When Should I be Checking It? \(NDSA, 2014\)](#)

Resources

- [HashMyFiles: Calculate MD5/SHA1/CRC32 hash of files](#)
- [Download DROID: file format identification tool - The National Archives](#)
- [DataAccessioner](#)
- [MD5summer](#)
- [Corz](#)
- [ExactFile | Making sure that what you hash is what you get.](#)
- [Levels of Digital Preservation](#)
- [Digital Readiness Levels – Digital Readiness Toolkit](#)

ON THE CALL TODAY

Questions? Contact the CMC: cmc@illinoisheartland.org



Barbera Scoby
CMC Cataloger



Dr. Pamela Thomas
Bibliographic Grant
Manager



Mary Cornell
CMC Cataloger



Eric McKinney
CMC Cataloging
Trainer



Kat Anderberg
Metadata Cataloger



Ryan Rafferty
Special Project Cataloger



Katie Roberts
Special Project Cataloger



Andrea Giosta
Special Project Cataloger