

Image Digitization: Best Practices and Training

ATALM Post-Conference SHN Workshop October 2016

Jeanine Nault, Lotus Norton-Wisla

Agenda (Day 2)

- 9:00-9:30 **Introduction**
- 9:30-10:00 **Technical Specifications**
- 10:00-11:00 **Hands-on Scanning**
- 11:00-11:30 **Discussion and Sharing**
- 11:30-12:00 **File Management and Quality Control**
- 1:00-1:30 **Metadata and Workflow**
- 1:30-1:45 **Resources**
- 1:45-2:00 **Questions and discussion**

How to start digitizing?

- In house
- Collaboration
- Outsourcing

Assessment - Factors to consider

- Size of collection
- Format
- Condition
- Time and Resources
- Budget

Good, Better, Best in Scanners

Lots of options! Consider:

- Results
- Resources available
- Format of materials



Good

\$100-200

Better

\$1500-1700

Best

\$2000-3000



Copy stand
with camera



Large
format



Slide
scanner

Technical Specifications and Scanning

Questions to Consider

- What is being digitized?
- Where are the files going?
- Where will they be stored?
- Who will create them?
- What guidelines will be followed?
- What are the technical specifications?

Quality in Digitizing

- Standards
- Versioning
- File Types
- File Size
- Resolution, Bit Depth, Color Profile
- Using Color Bars

Standards for Images

- Why follow standards?
- Where can I find them?

Standards

- Standards - always changing/developing
- FADGI
<http://www.digitizationguidelines.gov/guidelines/>
- Library of Congress Sustainable Formats
<http://www.digitalpreservation.gov/formats/intro/intro.shtml>
- Library of Congress, NARA, Universities

File Types and Sizes

- Recommended file types
- File size

File Formats

- Well supported?
- Open vs. proprietary?
- Quality vs. size
- Some common formats: TIFF, JPEG, DNG (RAW), JPEG 2000, PDF, PNG, GIF

File Copies - Master and Access

- Preservation Master
- Access Copy
- Web-ready derivative

File Size and Quality

Photographs

Performance Level:

	1 Star	2 Star	3 Star	4 Star
Master File Format	TIFF	TIFF	TIFF	TIFF
Access File Formats	All	All	All	All
Resolution	100 ppi	200 ppi	400 ppi	600 ppi ¹
Bit Depth	8	8	8 or 16	16
Color Space	Grey Gamma 2.2, SRGB, Adobe 1998, ECIRGBv2	Grey Gamma 2.2, Adobe 1998, ECIRGBv2	Adobe 1998, ECIRGBv2	Adobe 1998, ECIRGBv2
Color	Greyscale or Color	Greyscale or Color	Color	Color

Resolution

- Resolution: The number of pixels in each dimension that can be displayed - the density of pixels in the image.
- PPI: pixels per inch (DPI = dots per inch)
 - (300 ppi, 400ppi, 600ppi)

Preservation master: 4000-8000 pixels on long edge



The image to the left is from an image with nearly 3000 pixels along the long edge. The image on the right is from the same image with roughly 300 pixels along the long edge.

0%	255
10%	229
20%	203
30%	179
40%	159
50%	128
60%	102
70%	76
80%	52
90%	26
100%	0

Verso	
1.0	
2000	
1.5	
1350	
2.0	
1000	
2.5	800
3.0	675
3.5	575
4.0	500
4.5	450
5.0	400
5.5	365
6.0	335
6.5	310
7.0	290
7.5	270
8.0	250
8.5	240
9.0	225
9.5	215
10	
200+	

1	3000
1.5	2000
2.0	1500
2.5	1200
3.0	1000
3.5	875
4.0	750
4.5	675
5.0	600
5.5	550
6.0	500
6.5	475
7.0	450
7.5	400
8.0	375
8.5	375
9.0	350
9.5	325
10	300

1	6000
1.5	4000
2.0	3000
2.5	2400
3.0	2000
3.5	1750
4.0	1500
4.5	1350
5.0	1200
5.5	1100
6.0	1000
6.5	925
7.0	900
7.5	800
8.0	750
8.5	725
9.0	675
9.5	650
10	600

1	5000
1.5	3500
2.0	2500
2.5	2000
3.0	1675
3.5	1450
4.0	1250
4.5	1125
5.0	1000
5.5	925
6.0	850
6.5	775
7.0	725
7.5	675
8.0	625
8.5	600
9.0	575
9.5	550
10	500

1	4000
1.5	3000
2.0	2000
2.5	1600
3.0	1350
3.5	1150
4	1000
4.5	900
5.0	800
5.5	750
6.0	675
6.5	625
7.0	575
7.5	550
8.0	500
8.5	475
9.0	450
9.5	425
10	400

Bit Depth

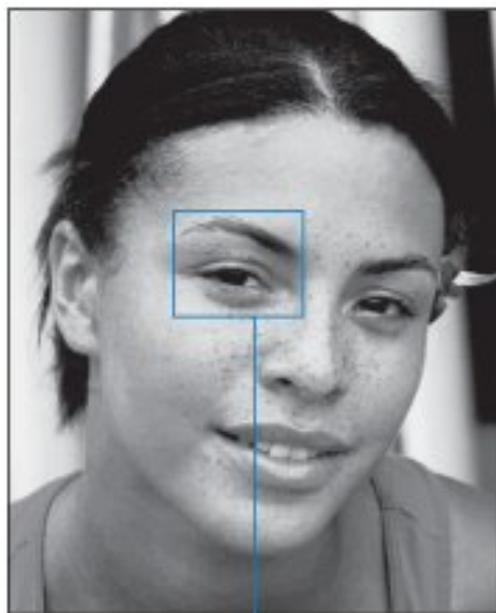
- Bit depth: the color information stored in an image. (The higher the bit depth, the more colors an image can store.)

2 bit: $2^1 = 4$ values

8 bit: $2^8 = 256$ colors

24 bit: $2^{24} = 16,000,000+$ colors

48 bit: $2^{48} = 3,000,000,000$ colors



1 bit
2 possible values



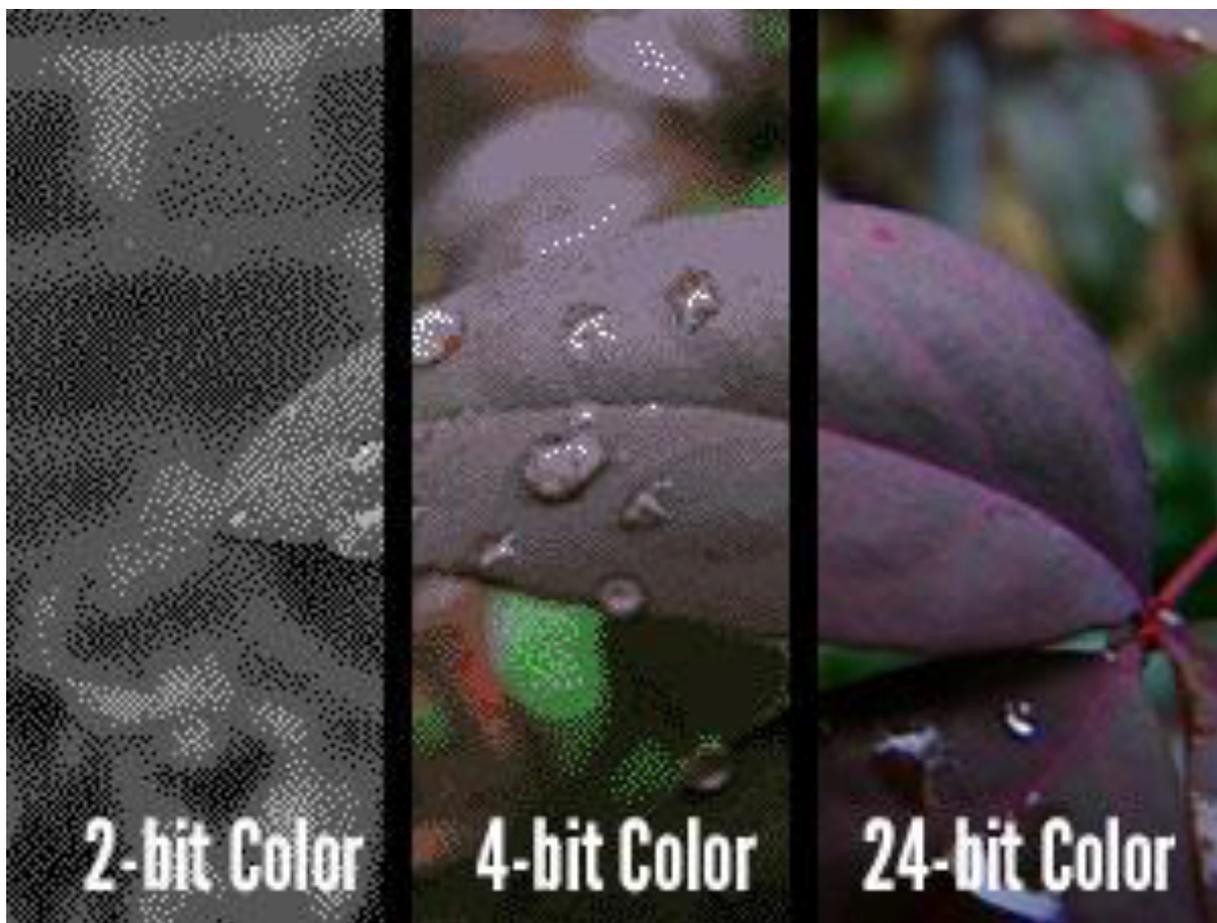
2 bits
4 possible values



4 bits
16 possible values



8 bits
256 possible values



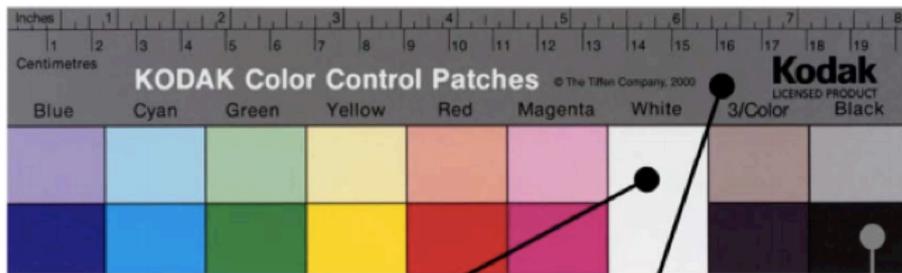
Color Profile and Tone

Color Profile: How colors combine over channels - each type of mode has a different number of channels.

Adobe RGB (1998) or Adobe sRGB

Tone: Range of values in an image - you want the whites not too bright and the blacks not too dark. Each pixel has a value of 0 - 255.

Using Color Bars in Scanning



		Neutralized White Point	Neutralized Mid Point*	Neutralized Black Point
Color Patch/Area		White	Gray Background	Single Color Black
Aimpoint	RGB Levels	237-237-237	102-102-102	23-23-23
	% Black	7%	60%	91%
Acceptable Range for Aimpoint	RGB Level	233 to 241	98 to 106	19 to 27
	% Black	5% to 9%	58% to 62%	89% to 93%

*Aimpoint for mid point (MP) to be calculated from actual values for white point (WP) and black point (BP) using the following formula: $MP = WP - 0.63(WP - BP)$

Hands-On Scanning Activity

Needs for Scanning

- Equipment:
 - Scanner
 - Computer
- Software:
 - Scanning software
 - Editing software
- Storage

Handling While Scanning

- Assess the physical condition of collections/ items to be scanned
- Wipe down scanner with lint-free cloth
- Clean, dry hands
- Hold by edges, carefully place and remove from scanner

File Edit View Tools Document Help

New Scan New Fax Forward as Fax Forward as E-mail Save as...

Scan

Fax

Scan

Date	File Name	File Type	Size	Source
5/15/201...	Welcome Scan	.jpg	504.3 KB	Windows Fax and Scan Team



Fax

Scan

New Scan



Scanner: WIA CanoScan LiDE 110

Change...

Profile: Photo (Default) ▾

Source: Flatbed ▾

Paper size: ▾

Color format: Color ▾

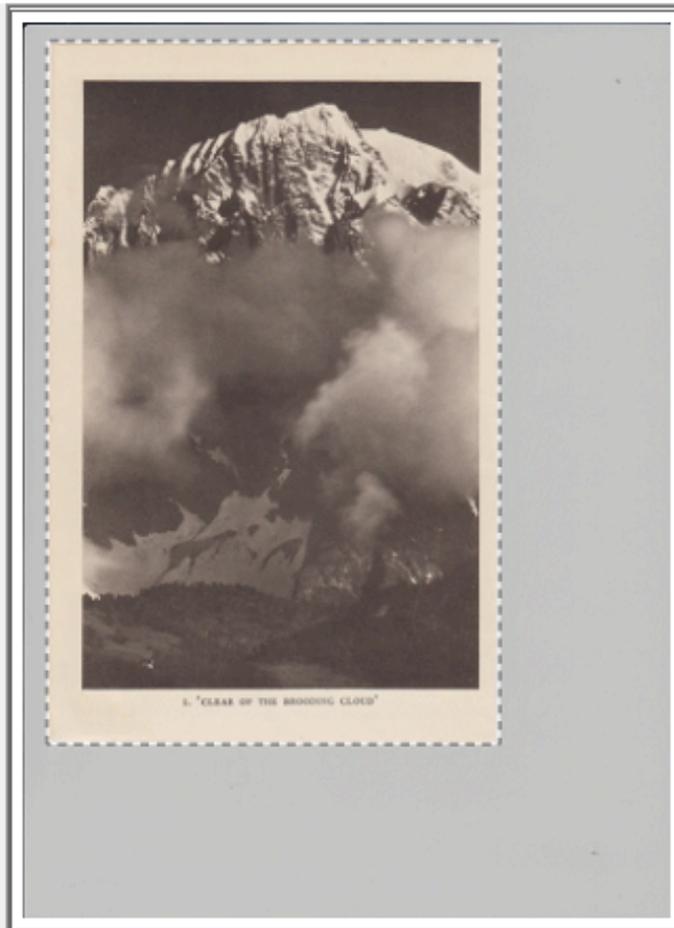
File type: JPG (JPG File) ▾

Resolution (DPI): 600

Brightness:  0

Contrast:  0

Preview or scan images as separate files



Preview

Scan

Cancel

Windows Fax and Scan

Edit View Tools Document Help

New Scan

New Fax



Forward as Fax



Forward as E-mail



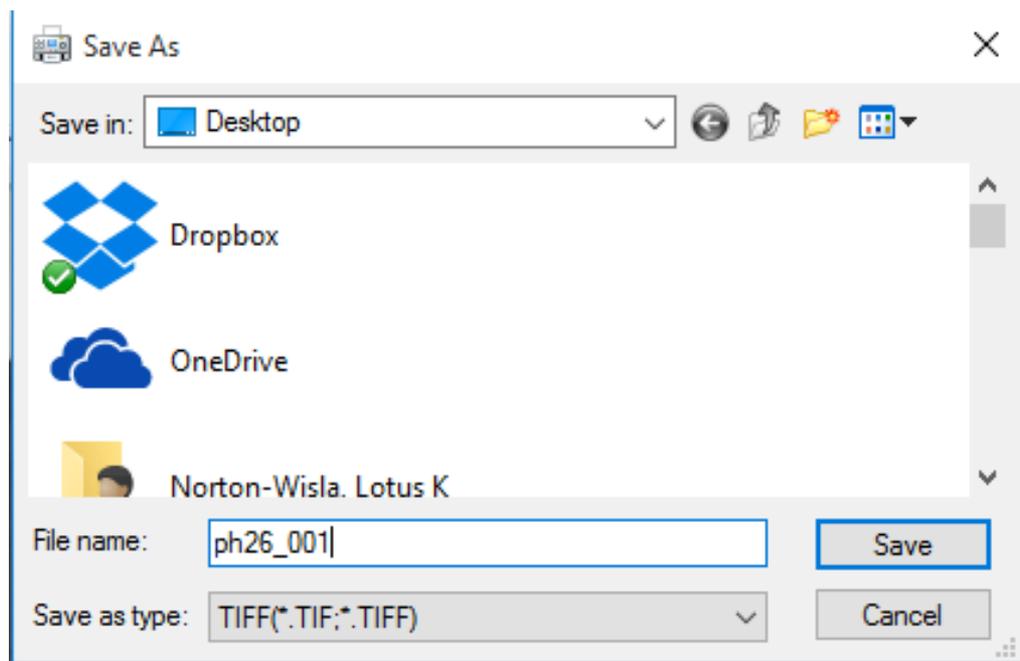
Save as...



Scan

Dat...	File Name	File Type	Size	So
5/15/201...	Welcome Scan	.jpg	504.3 KB	Wi
9/26/201...		.jpg	583.2 KB	WI

- View...
- Delete
- Rename...
- Send To >
 - Mail Recipient...
 - Fax...
- Move to Folder...
- Save As...
- Print...
- Zoom >

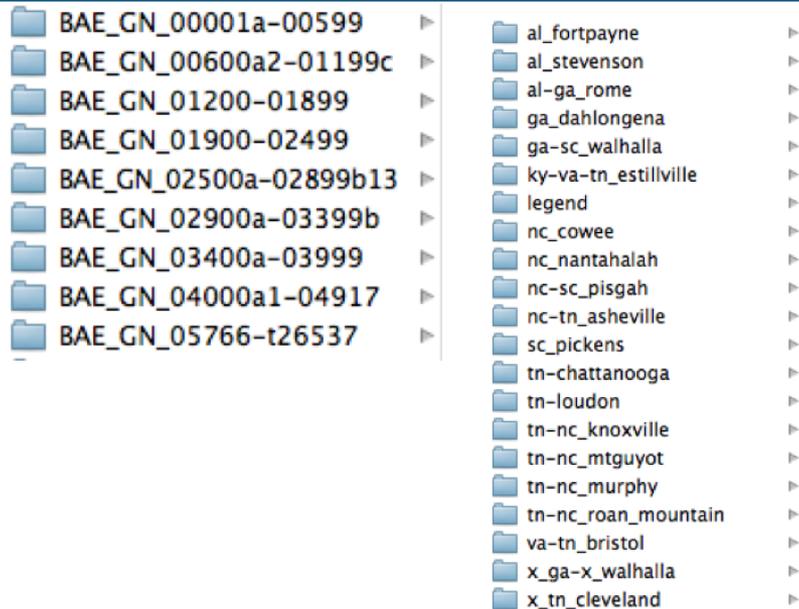


Activity Goals

- Groups: 4 groups of 3+
 - Goal: Each person scan **one** item
 - Identify any preservation issues
 - Find best resolution
 - Scan using color bar and settings
- Save as: **GroupX_01**
(Group1, Group2, or Group3)

File Management and Quality Control

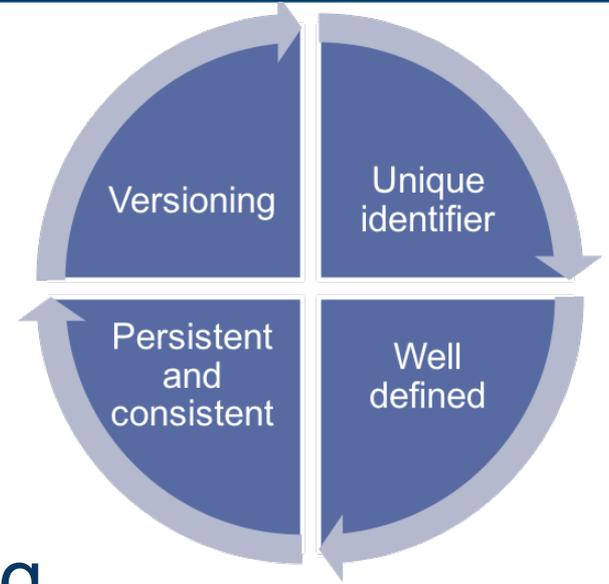
Folder Structure



- Master files
- Access copies
- Publication copies
- etc.

File Naming

- mary and joe at party.jpg
- Mary_joe_party.jpg
- Mary_joe_party_2015.jpg
- mary_joe_nye_party_2015.jpg



File Naming - Examples from NAA

- Prefix and suffix additions
- Examples of filename variations:
 - “V” = verso, backside of image
 - “x01” = numbering assigned during digitization
 - “ntbk” = manuscript is or contains bound notebook
 - “front_cover,” “back_cover,” “title_page”

Analog Collection Name	Digital Surrogate <u>File</u> naming Convention	Example
NAA MS 385	<u>msnumber_number.tif</u> ¹	385_002.tif
Notecard within NAA MS 385	<u>msnumber_notecard.tif</u>	385_notecard_1.tif (use number if sequence)
Notebook within NAA MS 385	<u>msnumber_ntbk_number.tif</u>	385_ntbk_001.tif If multiple notebooks: 385_ntbk_1_001.tif; 385_ntbk_2_001.tif
Kinship chart or other identifiable material within NAA MS 385	<u>msnumber_kinship_chart.tif</u>	351_kinshipchart_no_1_ver_2.tif ² ;
Note within NAA MS 385	<u>msnumber_front_note.tif</u> <u>msnumber_end_note.tif</u> <u>msnumber_pgnumber_note.tif</u>	385_front_note_01.tif, 385_end_note_01.tif, 385_01_note.tif

Quality Control

- Two or more step process
- Design for many stages of your workflow
- Develop from beginning, include Quality Control into your policies
- Sample large projects

Quality Control for Images

- Storage and organization
- Integrity
- Adherence to decided technical specifications
- Metadata
- Visual and inspection - digital artifacts, scanning mistakes



GUIDE TO QUALITY CONTROL AND QUALITY CHECKLISTS

What is Quality Control?

Digitization can be costly, take time, and mean extensive handling of original (and sometimes fragile) materials. For these reasons, the goal of any digitization project should be to create high-quality master images, audio, or video files from which several derivative images can be created for access and other uses. *Quality control* (QC) is an important part of any digitization project. QC includes procedures and techniques to verify the quality, accuracy, and consistency of digital files. Quality control should be conducted throughout all phases of the digital conversion process to ensure the materials need to be digitized only once, then can be used and shared many times.

Who is Responsible for Quality Control?

In most workflows, QC is performed in a two-step process: the person doing the converting performs an initial quality check during the digitization process, and then a different individual performs a second review in a separate process. If a vendor is conducting the digitization, this process will be different; however, it will still involve multiple stages of QC.

What Elements are You Checking in Quality Control?

Before you start scanning images or converting a/v materials, you need to determine the technical specifications (e.g. resolution, image mode, sample and bit rate, file format and storage medium) to use for your electronic documents in order to ensure their quality is preserved for the long term. This is the information that you will compare when you do your QC work. You should also clearly define the specific defects that you find unacceptable in a digital file so you and your staff know when a file needs to be re-digitized. If checking file fixity, you must decide on what method will be used to create and verify checksums.

Metadata and Workflow

Donor Forms (Deed of Gift)

- Photos, documents (etc.)
- Digital files
- Transfer (physical, legal, intellectual)
- Access restrictions

Donor Forms (Deed of Gift), continued

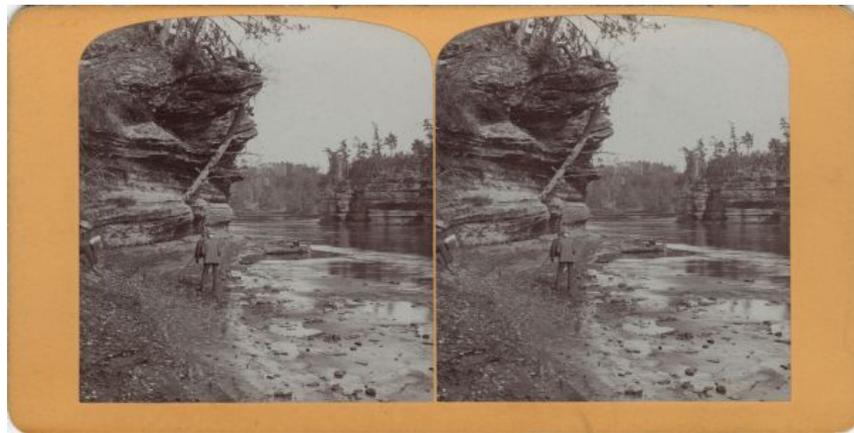
- Disposal or return of materials
- Communication with donors
- Get help if needed (admin, legal)
- Examples ([Karuk](#) Tribe, [USHMM](#))

Photographs in your Community

- Photos, documents
- Digital files
- Metadata, information, stories, memories

Metadata for Photographs

- Photograph type
- Family, community, land history
- Clothing and hairstyles
- Scenery, background
- Events, date information



Metadata for Photographs, continued

- Notch codes
- Photographer information
- Metadata written on or attached to photos
- Ask community members, elders for more information

Metadata Tools

- Spreadsheets
- Databases, CMSs
- Data dictionary - rules for how you enter metadata

Metadata Activity

- Fill out the metadata spreadsheet with your group for one item
- Postcards from the Nez Perce (Nee-Me-Poo) National Historic Trail

Project Planning - Can We Digitize?

- Equipment and software
- Physical Space
- Staff
- Digital storage needs
- Metadata
- Providing access

Project Management Tips

- Clear timeline
- Digitization Logs or Tracking Sheets
- Clear folder structure and file naming system
- Staff training

Project Planning and Workflow

What are the goals of the project?
What does a basic workflow look like?



Project Tracking

fishtank_update_current_master

Acc. No. Charles F. Hockett papers, 1934-2000, bulk 1940-1989
Title Unpublished 1969

Original Objects **JPEG Sent to SIRIS** No
Digital Images 8 **JPEG Attached to Record** No
Culturally_Sensitive No **JPEG in Reading Room** No
Items Not Scanned

Reason Oversize Bound Broken Fragile Other...

SIRIS Collection Record Yes **Hardware** PhaseOne IQ180
Item Level Records No

Missing Records

Notes associated audio text digitized under Arcadia project (IDS export in SOVA)

OCIO Backup **Archive Date Tivoli**

Master CD No **DAMS Ingest** Yes
Backup CD No **NAS backup** No

Scanned By Mig Dooley Johnson **Date** Apr 12, 2016
Proofed By Jeanine Nault **Date** Apr 25, 2016
PDF Available Yes **IDS Export** Complete
PDF linked in SIRIS No

100

ARCADIA DIGITIZATION PROJECT - DAMS ingest

updated 10/03/16 JN

Collection	Added - by	Number of Assets Ingested	Number of Assets Verified	Search Method in DAMS	Folder Created	Origination Format	Original Content Format	DAMS Level Metadata Applied
MS 2825	10/7/14 JN	154+2 PDFs	156		x		text	x
MS 3207	10/7/14 JN	19+1 PDF	20		x		text	x
MS 3353	10/7/14 JN	23+2 PDF	25		x		text	x
breslar_s2_taperec	8/27/14 SV	17+1 PDF	18		x		text	x
carlson_bk1	8/27/14 SV	65+1 PDF	66		x		text	x
carlson_ntbk10	8/27/14 SV	160+1 PDF	161		x		text	x
carlson_ntbk10b	8/27/14 SV	87+1 PDF	88		x		text	x
carlson_ntbk11	8/27/14 SV	118+1 PDF	119		x		text	x
carlson_ntbk12	8/27/14 SV	116+1 PDF	117		x		text	x
carlson_ntbk25	8/27/14 SV	261+1 PDF	262		x		text	x
carlson_ntbk26	8/28/14 SV	156+1 PDF	157		x		text	x
carlson_ntbk27	8/28/14 SV	68+1 PDF	69		x		text	x
carlson_ntbk28	8/28/14 SV	129+1 PDF	130		x		text	x
carlson_ntbk29	8/28/14 SV	96+1 PDF	97		x		text	x
harwood_s8_trans_taped_crmnies	8/28/14 SV	149+1 PDF	150		x		text	x
	11/12/14							
johnson_coeur-dalene	JN	87+1 PDF	88		x		text	x
	11/12/14							
johnson_s2_neahbay_ntes	JN	342+1 PDF	343		x		text	x
	10/17/14							
delaguna_sr	JN	107+46PDFs	153	"delaguna_sr"	x		text	x
	10/17/14							
delaguna_s2_1954_atna_rec	JN	38+1 PDF	39	"delaguna_s2_1954_atna_rec"	x		text	x

After Digitization: Editing

Reasons for Editing Software

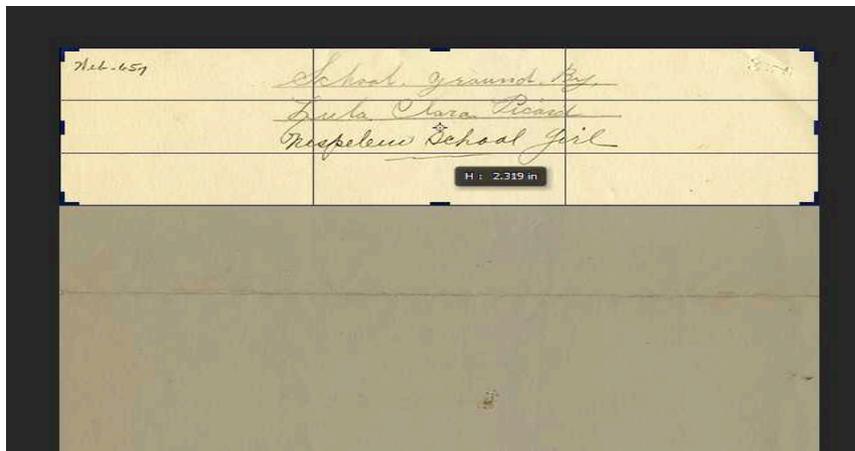
- Needed to edit images after scanning
- Needed to convert between common file formats
- Needed to examine images for QC
- If possible, find one piece of software that meets all your needs

Editing and Management Software

- Adobe Photoshop
- Adobe Lightroom
- Adobe Bridge
- GIMP
- IrfanView

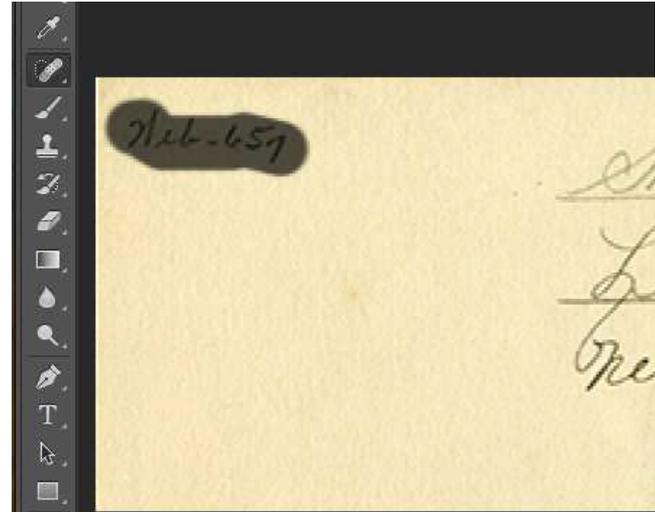
Editing and Processing Steps

- Rotating
- Exporting different types of files
- Changing resolution
- Cropping
- Automated batch actions



Editing and Processing Steps, continued

- Correcting errors
- Color correcting
- Adding text
- Branding
 - Watermarks



© 2014
je Patterson



Providing Access

Why Provide Access?

- Mission driven
- Reach newer, bigger audiences
- Expectation of public
- Feasibility
- Relevance
- Reduce handling, increase access to fragile collections

How to Provide Access?

- Who are your users?
- How will they access collections?
- How will you provide access?
- Open access, freely available? Or fee for service?
- Rights and Restrictions
- Copyright issues
- Staff, budget, time

Additional Resources

More Resources

- Sustainable Heritage Network
- Preservation Self Assessment Program
- FADGI
- Library of Congress
- Indigitization

Questions?

Thank you!

NaultJ@si.edu

lotus.norton-wisla@wsu.edu

support@sustainableheritagenetwork.org