Smithsonian
National Museum of Natural History

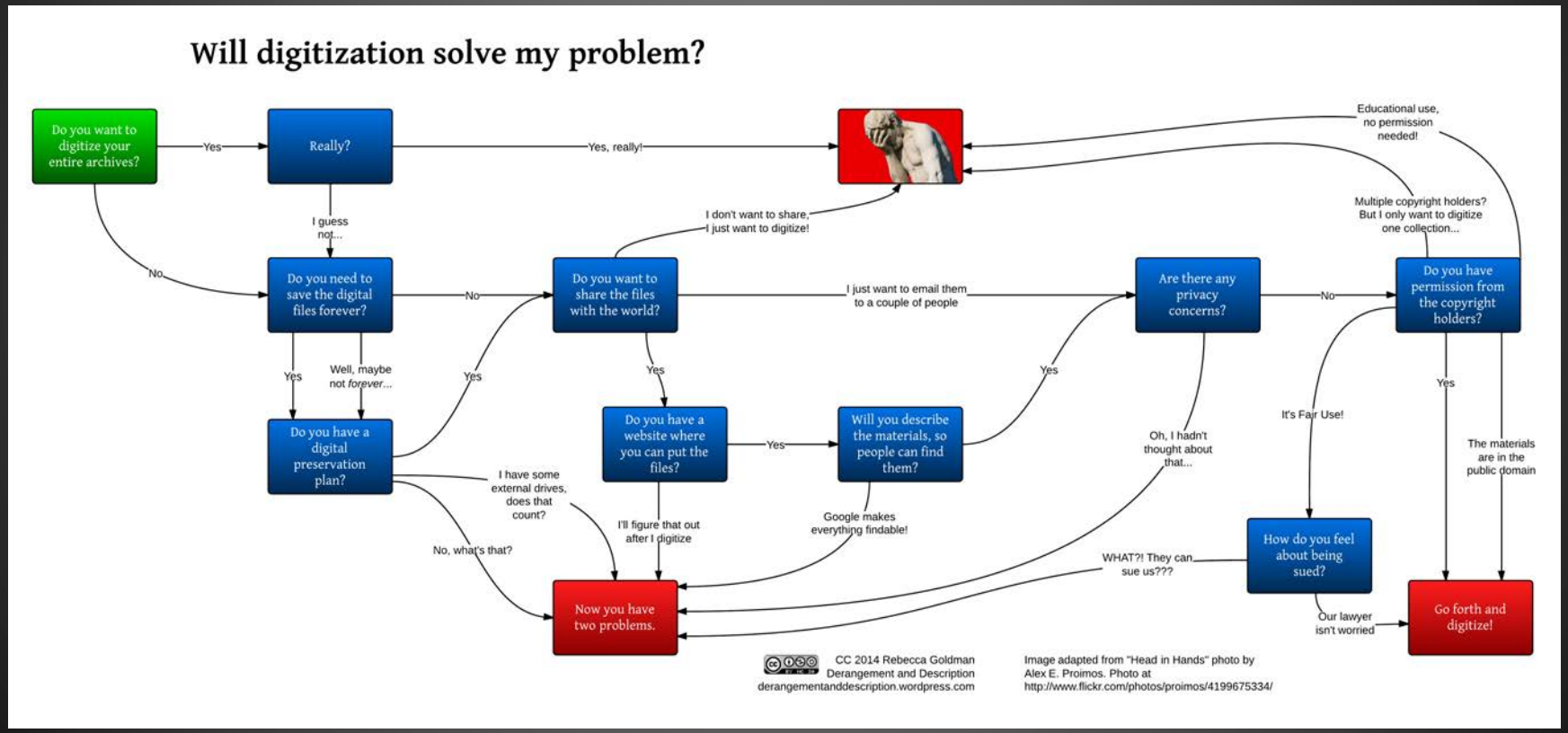
Bridging the Digital Divide: Basic Guidelines & Best Practices for digitization projects

Jeanine Nault
Digital Imaging Specialist
National Anthropological Archives
National Museum of Natural History
Smithsonian Institution
naultj@si.edu

Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators , April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

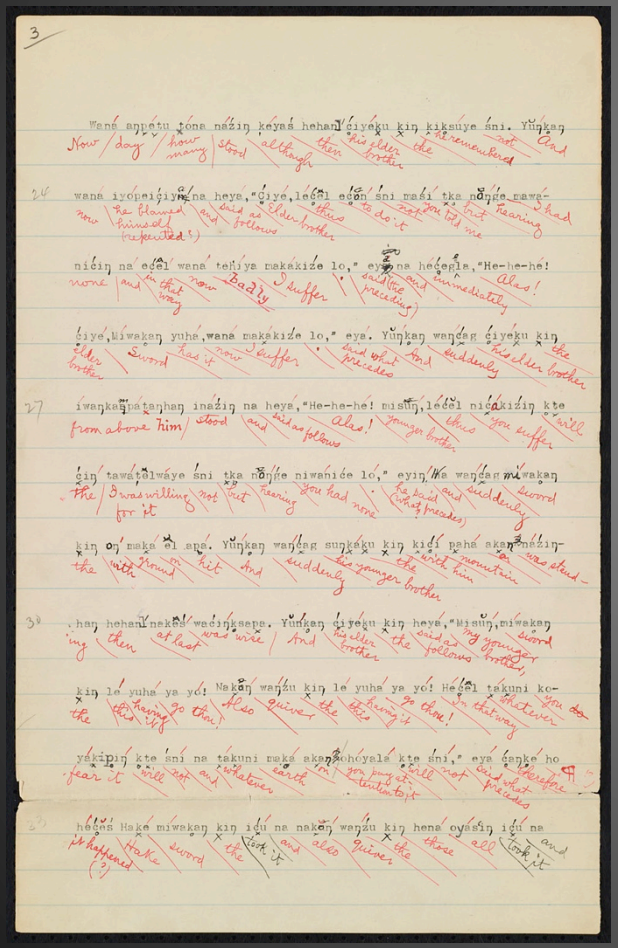
Smithsonian
National Museum of Natural History

Will digitization solve my problem?

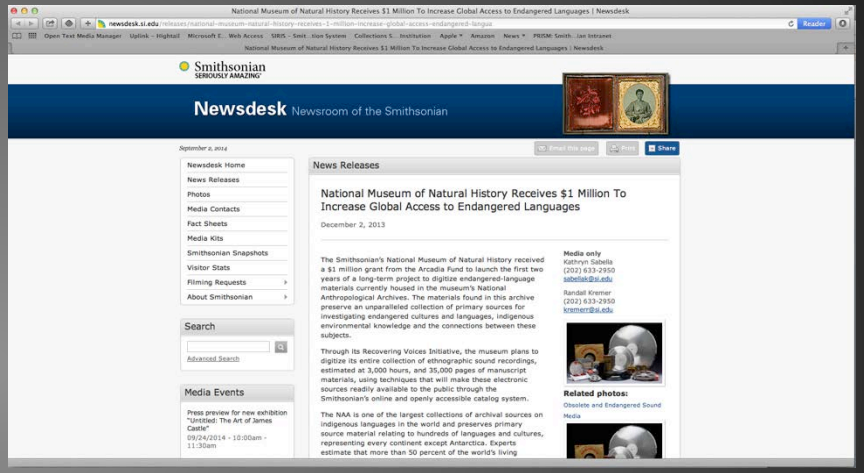


Smithsonian
National Museum of Natural History

National Anthropological Archives Arcadia project



[NAA MS 4800 [104]](), *National Anthropological Archives, Smithsonian Institution*

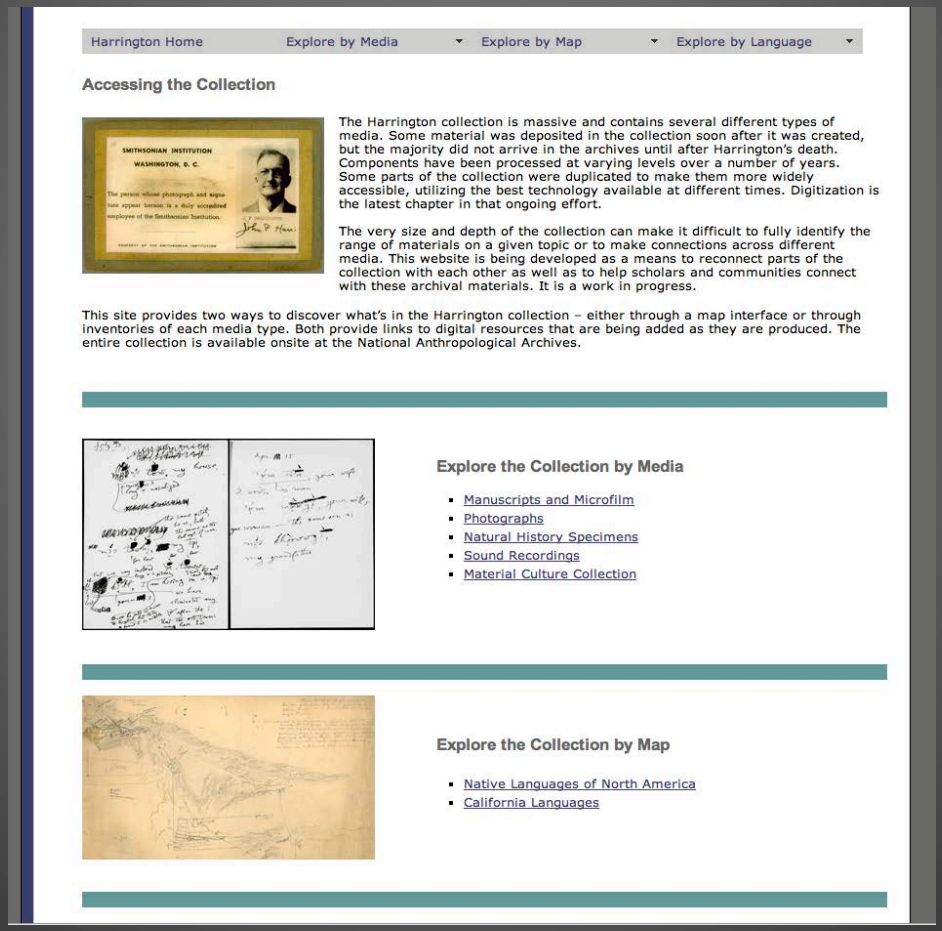


"Experts estimate that more than 50 percent of the world's living languages will cease to be spoken by the end of this century, and with them, the ability to confront worldwide environmental challenges will be diminished. The digitization project funded by Arcadia aims to conserve this knowledge, which is currently recorded on materials that are at high risk of degrading. The project team plans to create digital surrogates of voice recordings and paper documents in the NAA, and make them publicly available to support researchers and communities struggling to research, document and revitalize indigenous languages and cultures. Online access will make the material widely available for use without damage to the historic originals."

Smithsonian
National Museum of Natural History

John P. Harrington Collection web portal

http://anthropology.si.edu/naa/harrington/index.html



Smithsonian
National Museum of Natural History

Beyond slap & scan

Before beginning a digitization project, consider the following:

- WHY are we digitizing? (The lifecycle of the digital file – Where did it come from? Where is going?)
- WHAT are we digitizing? (Are there different types of materials with different needs?)
- HOW are we digitizing? (Equipment, software, people, time, money, space)
- WHO are we digitizing for? (Who will use these files? What will they do with it?)

Digitization projects are more than just slapping photos on a scanner

- Project planning and management
- Selection and Prioritization
- Technical Specifications
- Funding
- Preservation and Access
- Rights Issues
- To name just a few!

Smithsonian
National Museum of Natural History

Digitization Then and Now

Then

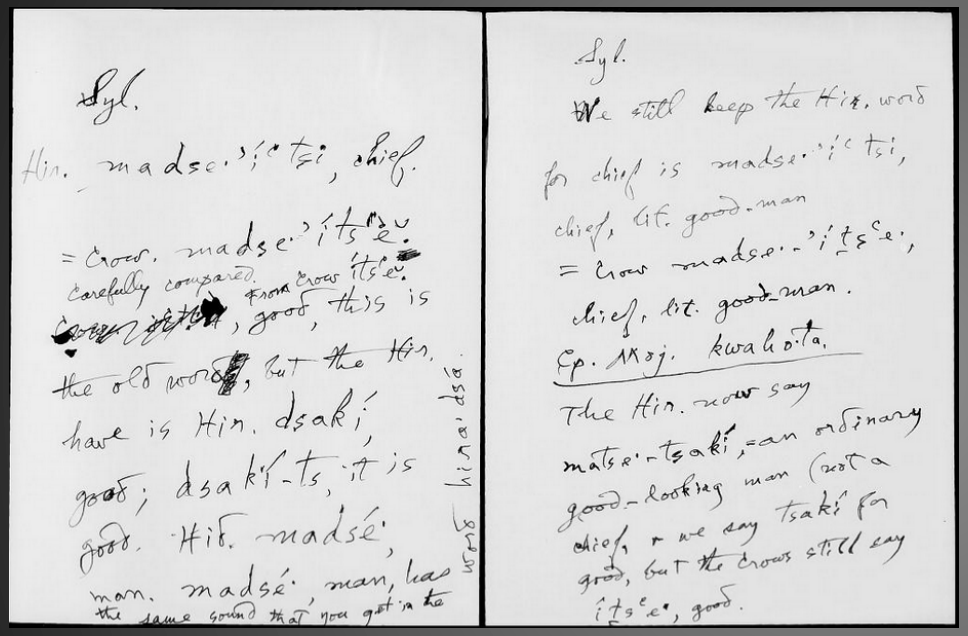
- Microfilm as nationally accepted standard for decades – cost effective way to preserve text materials
  - But what about graphic materials like paintings and photos?
- Copy negatives – expensive and two generations removed from original
  - Original object → New negative → Copy print

Now

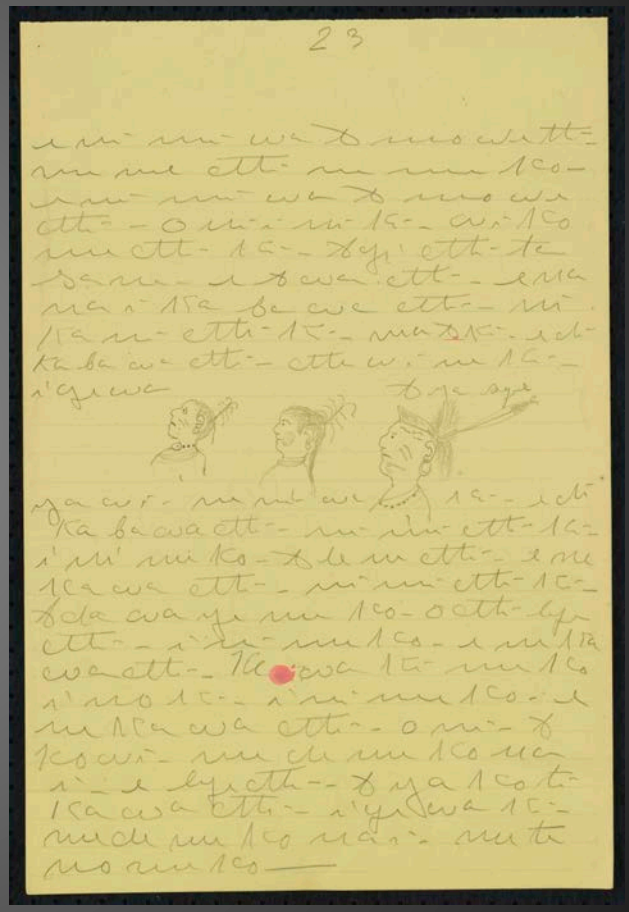
- Digital imaging and digitization allows for preservation of original object (less handling, high resolution file for detailed viewing) and increases access to collections

Smithsonian
National Museum of Natural History

Then and Now



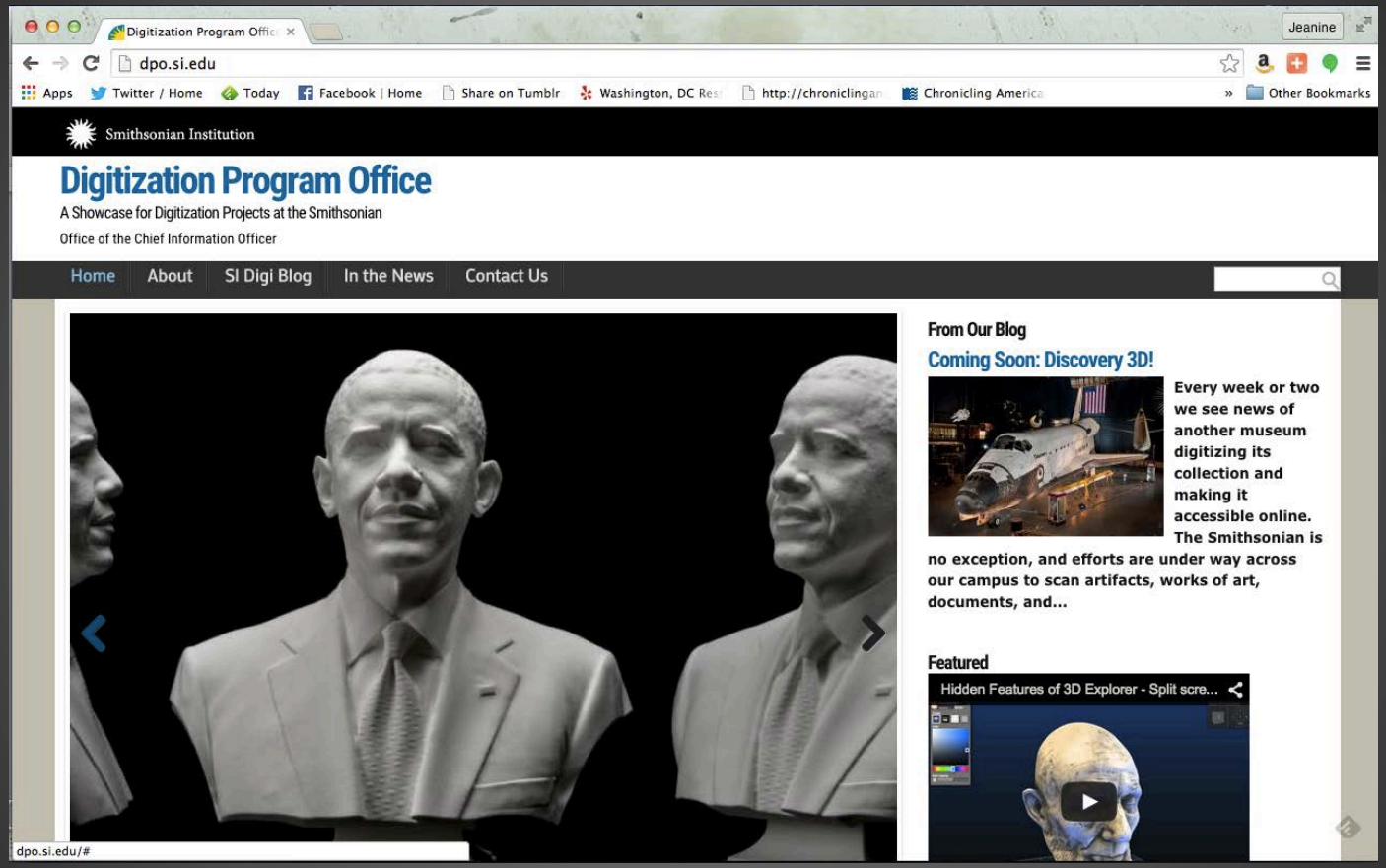
John Peabody Harrington papers: Mandan/ Hidatsa/Crow, 1950-1951
Image name: Accession #1976-95 [NMNH-Harrington_mf5_r14_0002]



NAA MS 2046, Fox text by Alfred Kiyana on Buffalo dance, undated

Smithsonian
National Museum of Natural History

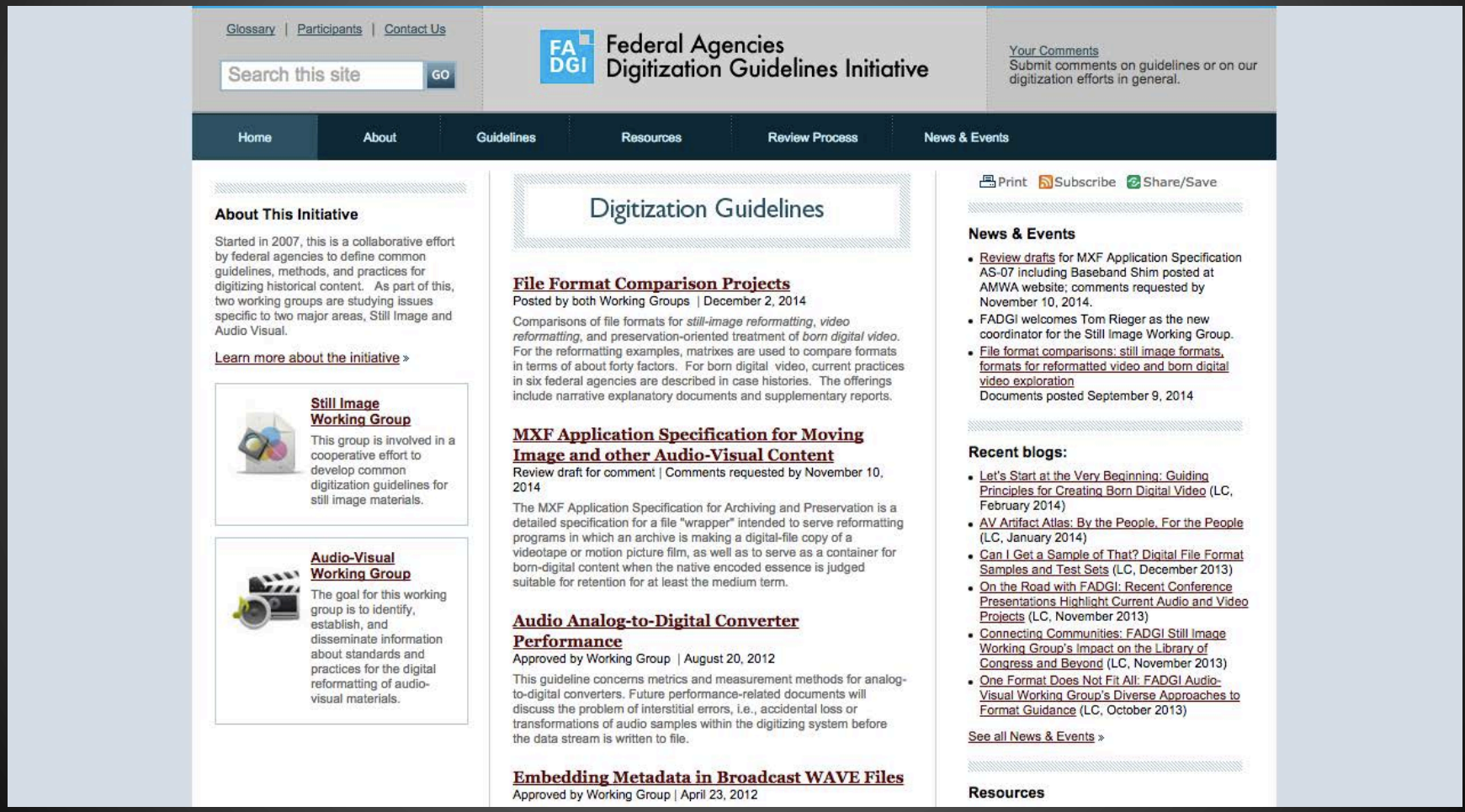
Digitization Now



- [Smithsonian Digitization Program Office Rapid Capture Digitization Process]

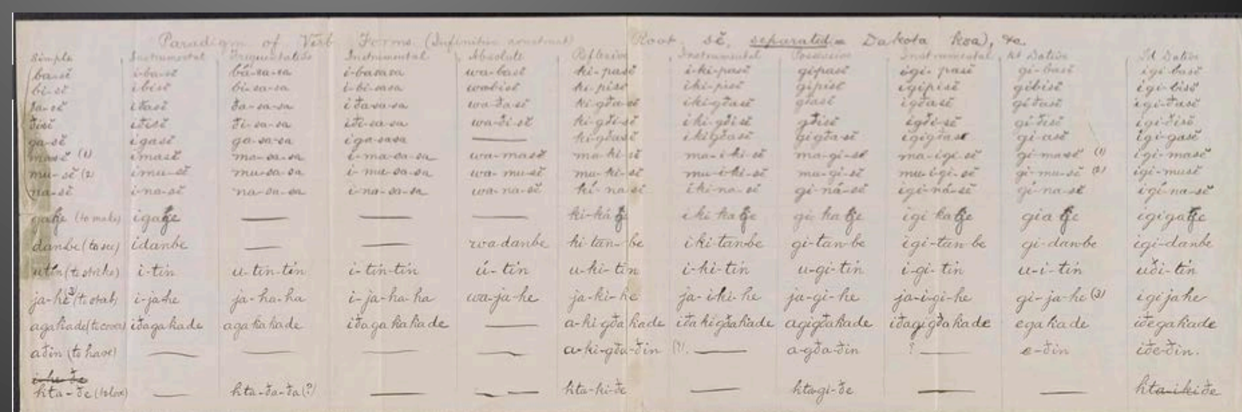
Smithsonian
National Museum of Natural History

FADGI: Federal Agencies Digitization Guidelines Initiative



Smithsonian
National Museum of Natural History

Materials: What are we working with?

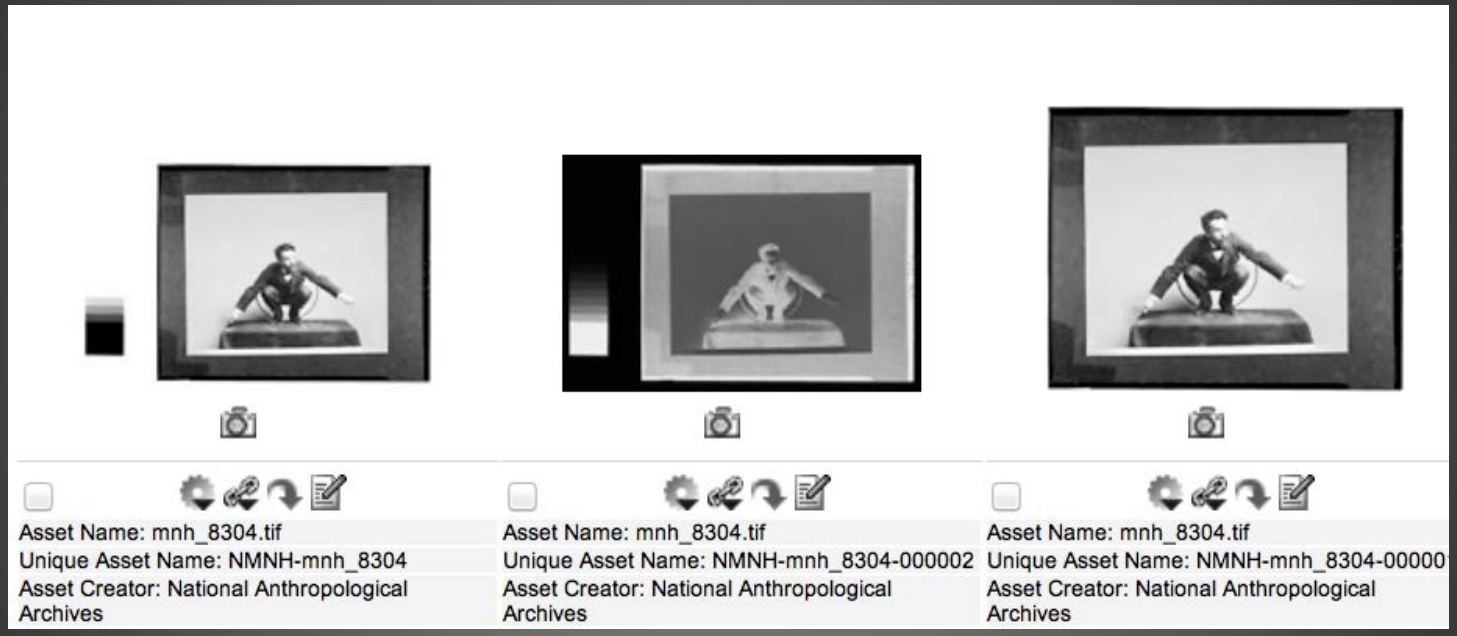


Paradigm of Verb Forms (Infinitive construct) Root sĕ, separated (in Dakota ksa), &c.

Simple	Instrumental	Frequentative	Instrumental	Absolute	Reflexive	Instrumental	Possessive	Instrumental	1st Dative	2d Dative
basĕ	i-ba-sĕ	bá-sa-sa	i-basasa	wa-basĕ	ki-pasĕ	i-ki-pasĕ	gipasĕ	igi-pasĕ	gi-basĕ	igi-basĕ
bi-sĕ	ibisĕ	bi-sa-sa	i-bi-sasa	wabisĕ	ki-pisĕ	iki-pisĕ	gipisĕ	igipisĕ	gibisĕ	igi-bisĕ
đa-sĕ	iđasĕ	đa-sa-sa	iđasa-sa	wa-đa-sĕ	ki-gđa-sĕ	iki-gđasĕ	gđasĕ	igđasĕ	giđasĕ	igi-đasĕ
đisĕ	iđisĕ	đi-sa-sa	iđi-sa-sa	wa-đi-sĕ	ki-gđi-sĕ	iki-gđisĕ	gđisĕ	igđi-sĕ	giđisĕ	igi-đisĕ
ga-sĕ	igasĕ	ga-sa-sa	iga-sasa	—	ki-gđasĕ	ikigđasĕ	gigđa-sĕ	igigđasĕ	gi-asĕ	igi-gasĕ
masĕ (1)	imasĕ	ma-sa-sa	i-ma-sa-sa	wa-masĕ	ma-ki-sĕ	ma-i-ki-sĕ	ma-gi-sĕ	ma-igi-sĕ	gi-masĕ (1)	igi-masĕ
mu-sĕ (2)	imu-sĕ	mu-sa-sa	i-mu-sa-sa	wa-mu-sĕ	mu-ki-sĕ	mu-i-ki-sĕ	mu-gi-sĕ	mu-igi-sĕ	gi-mu-sĕ (2)	igi-musĕ
na-sĕ	i-na-sĕ	na-sa-sa	i-na-sa-sa	wa-na-sĕ	ki-na-sĕ	iki-na-sĕ	gi-ná-sĕ	igi-ná-sĕ	gi-na-sĕ	igi-na-sĕ
gaҟe (to make)	igaҟe	—	—	—	ki-káҟe	iki kaҟe	gi kaҟe	igi kaҟe	gia ҟe	igigaҟe
danbe (to see)	idanbe	—	—	wa danbe	ki-tan-be	iki-tanbe	gi-tan-be	igi-tan-be	gi-danbe	igi-danbe
utin (to strike)	i-tin	u-tin-tin	i-tin-tin	ú-tin	u-ki-tin	i-ki-tin	u-gi-tin	i-gi-tin	u-i-tin	uđi-tin
ja-he (3) (to stab)	i-jahe	ja-ha-ha	i-ja-ha-ha	wa-ja-he	ja-ki-he	ja-iki-he	ja-gi-he	ja-igi-he	gi-ja-he (3)	igijahe
agakade (to cover)	iđagakade	agakakade	iđagakakade	—	a-ki-gđa kade	iđakigđakade	agigđakade	iđagigđakade	egakade	iđegakade
ađin (to have)	—	—	—	—	a-ki-gđa-đin	(?) —	a-gđa-đin	? —	e-đin	iđe-đin.
~~i-he-đe~~ hta-đe (to love)	—	hta-đa-đa (?)	—	—	hta-ki-đe	—	htagi-đe	—	—	hta-ikiđe

Smithsonian
National Museum of Natural History

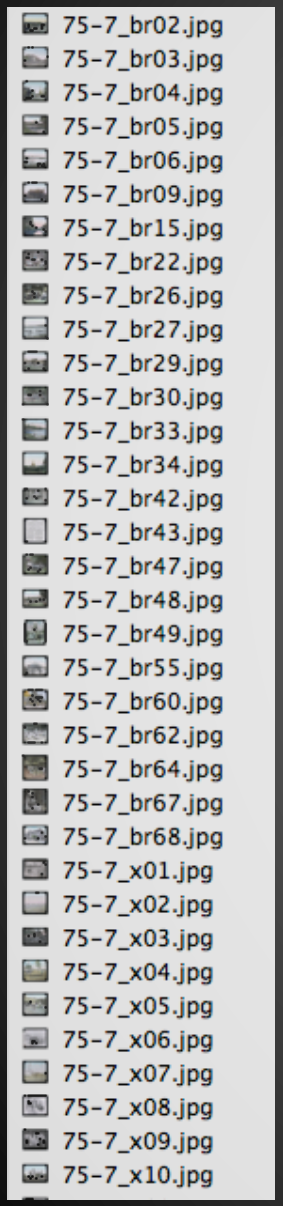
File Management and Organizing your Workflow



Above image: MNH 8304, Franz Boas posing for figure in USNM exhibit entitled "Hamats'a coming out of secret room" 1895 or before*, National Anthropological Archives, Smithsonian Institution*

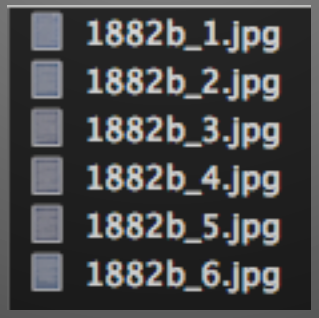
Smithsonian
National Museum of Natural History

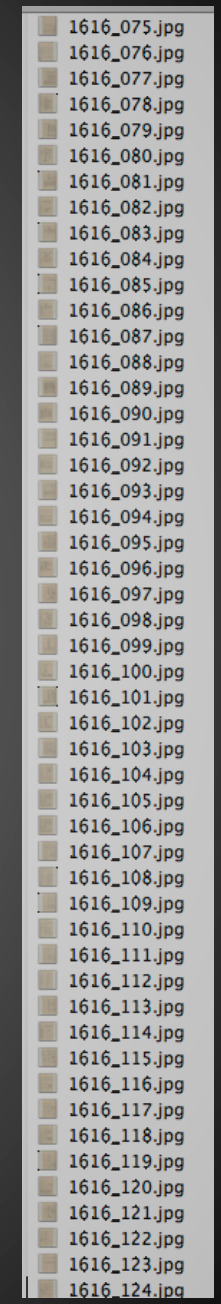
File Management and Organizing your Workflow



What are the Issues?

- filenaming
- formats/compatibility
- metadata
- Storage/use

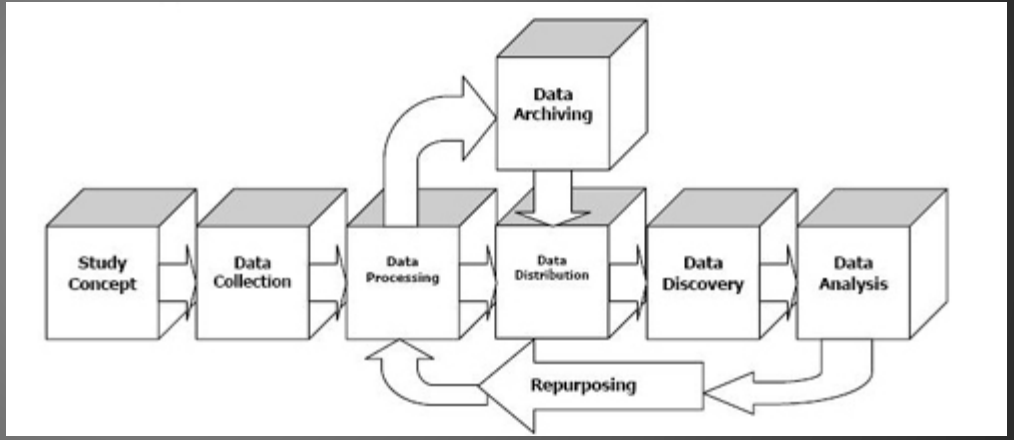




Smithsonian
National Museum of Natural History

File Management and Organizing your Workflow

- Project planning
- Project tracking
- Selection and Prioritization
- Image Capture
- Post-production and Quality Control
- Secure storage and migration



Source: DDI Structural Reform Group. "DDI Version 3.0 Conceptual Model." *DDI Alliance*. 2004. <http://www.icpsr.umich.edu/DDI/committee-info/Concept-Model-WD.pdf>.

Smithsonian
National Museum of Natural History

Project Planning: File Management and Organizing your Workflow

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

Smithsonian
National Museum of Natural History

Project Planning: File Management and Organizing your Workflow

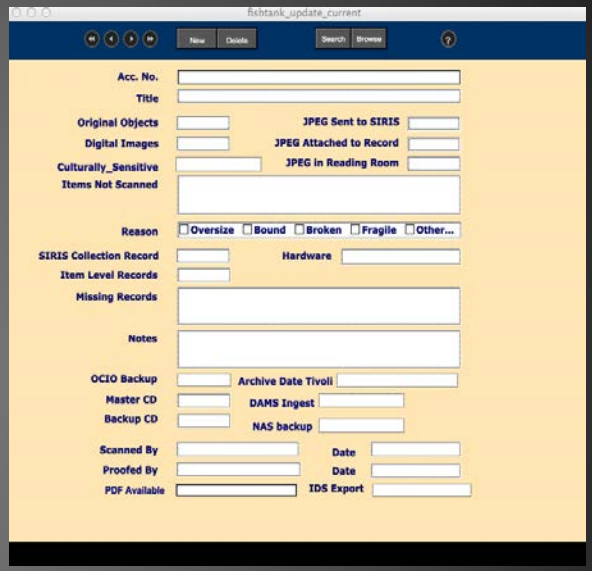
Sketch out a Workflow Diagram

- What are the specific materials you are going to digitize?
- Do you have enough storage space for the files?
- Begin to develop a filenaming schema – relate to collections naming as much as possible
- What metadata will you include?
- Lifecycle of the image
- Storage and migration

Smithsonian
National Museum of Natural History

Project Tracking: File Management and Organizing your Workflow

Language	NAA Local Control Number	Title (Siris Record)	Creator	Physical Description (size etc.)	Scanning Instructions	Observations made during preliminary review (box #)	Conservation Y/N	Digitization Request Date	# of pages for digitization	Digitization status	PDF complete
Menominee	NAA MS 2797	Texts	Michelson, Truman	13 pages	SELECTED FOR DIGITIZATION NOTE: the paper is very brittle, so for the purposes of conservation it would be ideal to scan all pages. If this is not possible, then follow these instructions: Please scan pp. 1-28. Also scan pp. 1-6 and 7-12.	Handwritten notes that appear to be of texts. Several dozen pages are in Menominee. Some pages are in English but it is not clear whether these are translations of the Menominee texts. The Menominee texts seems to be paginated in an odd way (or are out of order). The Menominee pages include some grammatical and lexical data as well. Perhaps we could digitize some texts. Conservation: paper is very brittle and breaks to the touch. Recommendation to Bcs, participants: Please review the digital surrogates against the physical manuscript and againts MS 2824 to determine whether the English texts in those records are the translations of the Menominee texts.	Y	20-Mar	41	complete	x
Maidu	NAA MS 646-b	Linguistic material 1877	Gatschet, Albert S. (Albert Samuel), 1832-1907 collector	84 pages	Please scan until page 1-33, 91-102, 129-131 for Bcs. However, for conservation purposes, I recommend the entire manuscript be scanned.	Substantial sentencial data, some vocabulary. I have selected this manuscript because our holdings of Konkow are modest, yet Chico Maidu is often considered to be the same language as Konkow (Golla 2011:138). Conservation: the ink seems to be fading or at least it is very light. Scanning may be recommended for conservation purposes.		26-Mar	45	complete	x
Maidu	NAA MS 646-c	Supplementary Maidu vocabulary and sentences	ANONYMOUS	3 pages	Please scan all 3 pages	Konkan dialect. Limited data.	N	26-Mar	5	complete	x
Hupa	NAA MS 1442	Hupa vocabulary December 1888-January 1889	Curtin, Jeremiah	107 pages	Please scan: p 82-83; 88-96; 109-125; 132-133.		N	26-Mar	28	complete	x
Hupa	NAA MS 934	Weitspek (Pohlik Klamath) and Hoopah dictionaries and ethnographic notes 1852	Gibbs, George	41 pages autograph document signed	Please scan entire notebook marked Shelf 71			2-Apr	29	complete	x
Karok	123	Karok Medicine Formulas	Harrington, John Peabody 1884-1961		please scan first 11 page (marked) of 123-1		Y	2-Apr	11	complete	x
Karok	130	Typescript of MS "[Ethnobotany and Ethnozoology of the Karuk Indians of California" (former BAE MS#4553)	Harrington, John Peabody 1884-1961		please scan marked pages (10 pages) in 130-3 and marked pages (11 pages) in 130-2			2-Apr	21	complete	x
Karok	122	Karok Stories and Myths	Harrington, John Peabody 1884-1961		Mockingbird and Robin story, Box 122, Folder 4, 11 pages separated by paper slip.		N	28-Mar	11	complete	x
Fernandeño	563-1	Linguistic and Ethnographic Field Notes	Harrington, John Peabody 1884-1962		Please scan all 12 handwritten letter-size pages and the first 10 typewritten paper slips.	The first slip on the data slip pack is somewhat damaged.	Y	28-Mar	17	complete	x

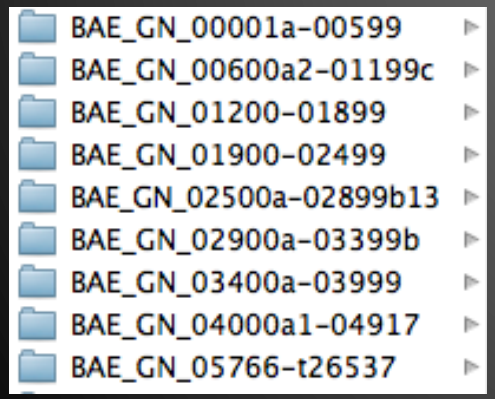


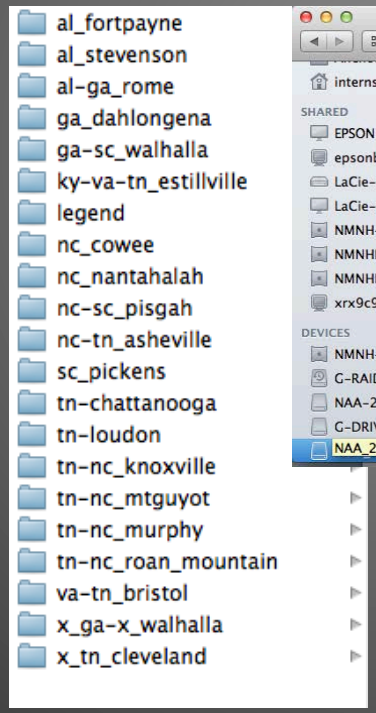
Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators , April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

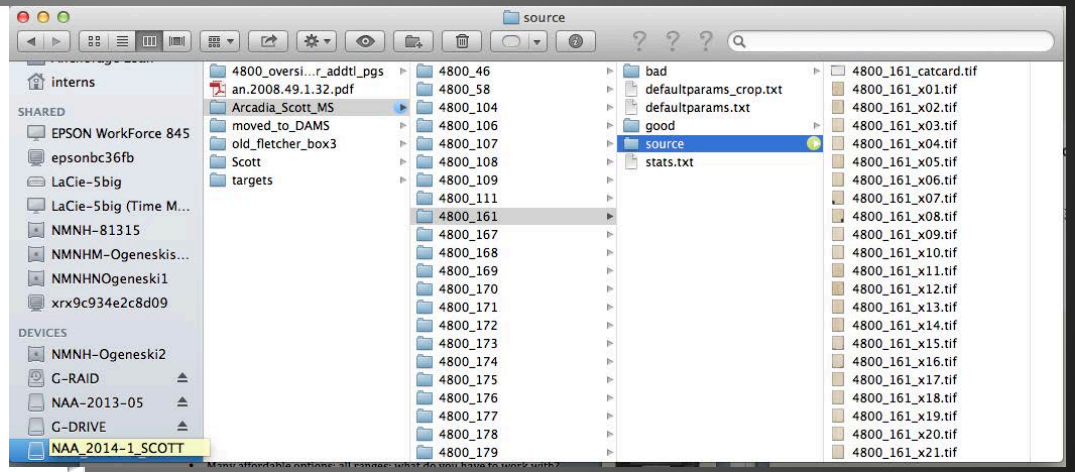
Smithsonian
National Museum of Natural History

Project Tracking: File Management and Organizing your Workflow

Folders, Folders, Folders: How you name them matters!





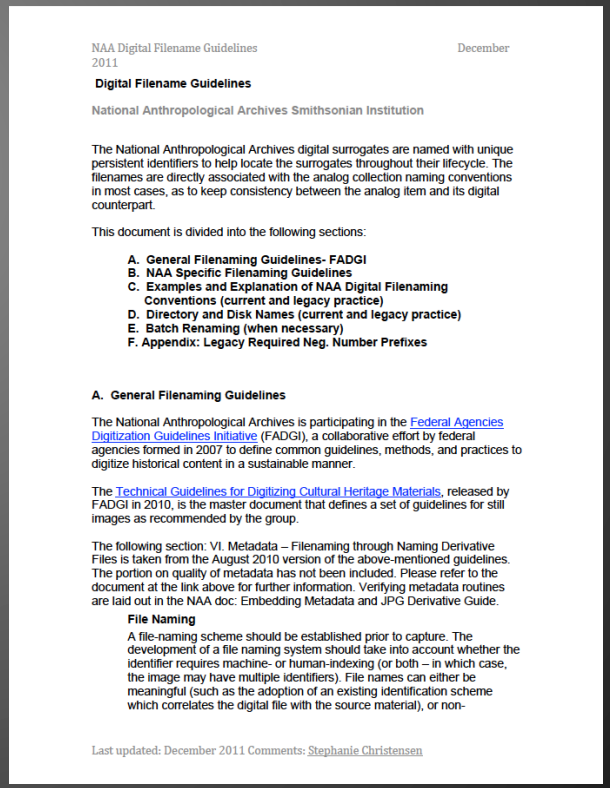


Smithsonian
National Museum of Natural History

Pre-Digitization Work: File Management and Organizing your Workflow

Filenaming, Filenaming, Filenaming, Filenaming, Filenaming!

- Unique identifier; well defined
- Persistent and consistent
- Versioning



NAA Digital Filename Guidelines December 2011

Digital Filename Guidelines

National Anthropological Archives Smithsonian Institution

The National Anthropological Archives digital surrogates are named with unique persistent identifiers to help locate the surrogates throughout their lifecycle. The filenames are directly associated with the analog collection naming conventions in most cases, as to keep consistency between the analog item and its digital counterpart.

This document is divided into the following sections:

A. General Filenaming Guidelines- FADGI
B. NAA Specific Filenaming Guidelines
C. Examples and Explanation of NAA Digital Filenaming Conventions (current and legacy practice)
D. Directory and Disk Names (current and legacy practice)
E. Batch Renaming (when necessary)
F. Appendix: Legacy Required Neg. Number Prefixes

A. General Filenaming Guidelines

The National Anthropological Archives is participating in the Federal Agencies Digitization Guidelines Initiative (FADGI), a collaborative effort by federal agencies formed in 2007 to define common guidelines, methods, and practices to digitize historical content in a sustainable manner.

The Technical Guidelines for Digitizing Cultural Heritage Materials, released by FADGI in 2010, is the master document that defines a set of guidelines for still images as recommended by the group.

The following section: VI. Metadata – Filenaming through Naming Derivative Files is taken from the August 2010 version of the above-mentioned guidelines. The portion on quality of metadata has not been included. Please refer to the document at the link above for further information. Verifying metadata routines are laid out in the NAA doc: Embedding Metadata and JPG Derivative Guide.

File Naming

A file-naming scheme should be established prior to capture. The development of a file naming system should take into account whether the identifier requires machine- or human-indexing (or both – in which case, the image may have multiple identifiers). File names can either be meaningful (such as the adoption of an existing identification scheme which correlates the digital file with the source material), or non-

Last updated: December 2011 Comments: Stephanie Christensen

Smithsonian
National Museum of Natural History

Pre-Digitization Work: File Management and Organizing your Workflow

Analog Collection Name	Digital Surrogate Filenaming Convention	Example
NAA MS 385	Msnumber_number.tif[1]	385_002.tif
Notecard within NAA MS 385	Mnsumber_notecard.tif	385_notecard_1.tif (use number if sequence)
Notebook within NAA MS 385	Msnumber_ntbk_number.tif	385_ntbk_001.tif (multiple notebooks: 385_ntbk_1_001.tif, 385_ntbk_2_001.tif, etc
Multiple folders of NAA MS 4117-a	MSnumber_foldernumber_number.tif	4117-a_f1_001.tif, 4117-a_f2_001.tif, 4117-a_f3_001.tif, 4117-a_f4_001.tif
Multiple volumes of NAA MS 1795a	Msnumber_volnumber_number.tif	1795a_vol1_001.tif, 1795a_vol2_001.tif, 1795a_vol3_001.tif, 1795a_vol4_001.tif, 1795a_vol4_drawings_001.tif
Kinship chart or other identifiable materials within NAA MS 385	Msnumber_kinship_chart.tif	351_kinshipchart_no_1_ver_2.tif[2]

Smithsonian
National Museum of Natural History

Pre-Digitization Work: File Management and Organizing your Workflow

Note within NAA MS 385	Msnumber_note.tif	385_note.tif; 385_note_1.tif; 385_note_2.tif
Drawing within NAA MS 385	Msnumber_drawing.tif	385_drawing.tif*
Map within NAA MS 385	Msnumber_map.tif or msnumber_map_identifiableinfo_tif.	385_map_001.tif; 385_map_east_facade.tif
Catalog cards	Msnumber_catcard.tif	385_catcard.tif; 385_catcard_v.tif[3]
MS 176,622	These types of numbering are also used with manuscripts and items that may contain INV numbers	0851501.tif
MS 166,032	Directory folder is 166032	Folder uses INV numbers 08704000
Ledger art	NMNHnaa_la_Anon_Plains_08531400_post.tif	File given out to public—drop prefix and suffix—distribute: 08513400.tif
Ledger art	May use INV numbers	**11005600.tif 08624400.tif**
Ledger art	NMNHnaa_la_ms4653_08691300_post	Contained within MS 4653—drop prefix and suffix; distribute: 08691300.tif

Smithsonian
National Museum of Natural History

Pre-Digitization Work: File Management and Organizing your Workflow

Filenaming, Filenaming, Filenaming, Filenaming, Filenaming, Filenaming,...

- Prefix and suffix additions
- Examples of filename variations:
  - "v" = verso, the backside of a page
  - "x01" = numbering assigned during digitization
- "ntbk" – manuscript is a bound notebook
- "front_cover," "back_cover" "title_page"

MS 4490	
Main file to distribute/ combines all other files	4490.tif
These files illustrate Left, right, mid section and additional "detail views" that may have been shot	4490_1_detail.tif 4490_2_detail.tif 4490_3_detail.tif 4490_4_detail.tif 4490_5_detail.tif 4490_6_detail.tif msnumber_left_1.tif; msnumber_right_1.tif

Smithsonian
National Museum of Natural History

Actual Digitization: File Management and Organizing your Workflow

File Format

.tiff .jpg .mov .wav .mp3 .pdf

Are you saving in a file format that is compatible in the long term?

Hi- bit or low bit depth; Color space

Smithsonian
National Museum of Natural History

Actual Digitization: File Management and Organizing your Workflow

Secure Storage and Migration

- Can you access your files easily?
- Do you understand your filenaming months later after you've been away from the project?
- Do you have multiple copies in separate locations?
- What is your storage media?
- Do you have a long- term plan for the security of you data?
- If so you are on your way!

Smithsonian
National Museum of Natural History

File Management and Organizing your Workflow

Additional Resources

- Federal Agencies Digitization Guidelines Initiative
- University of Cambridge Library – Data Management, Choosing File Formats
- MIT Libraries: Data Management and Publishing

Smithsonian
National Museum of Natural History

Digitization Project Management Selection and Prioritization, A Case Study: Breath of Life Archival Institute: 2013, 2015

Jeanine Nault
Digital Imaging Specialist
National Anthropological Archives
National Museum of Natural History
Smithsonian Institution
naultj@si.edu

Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators , April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

Smithsonian
National Museum of Natural History

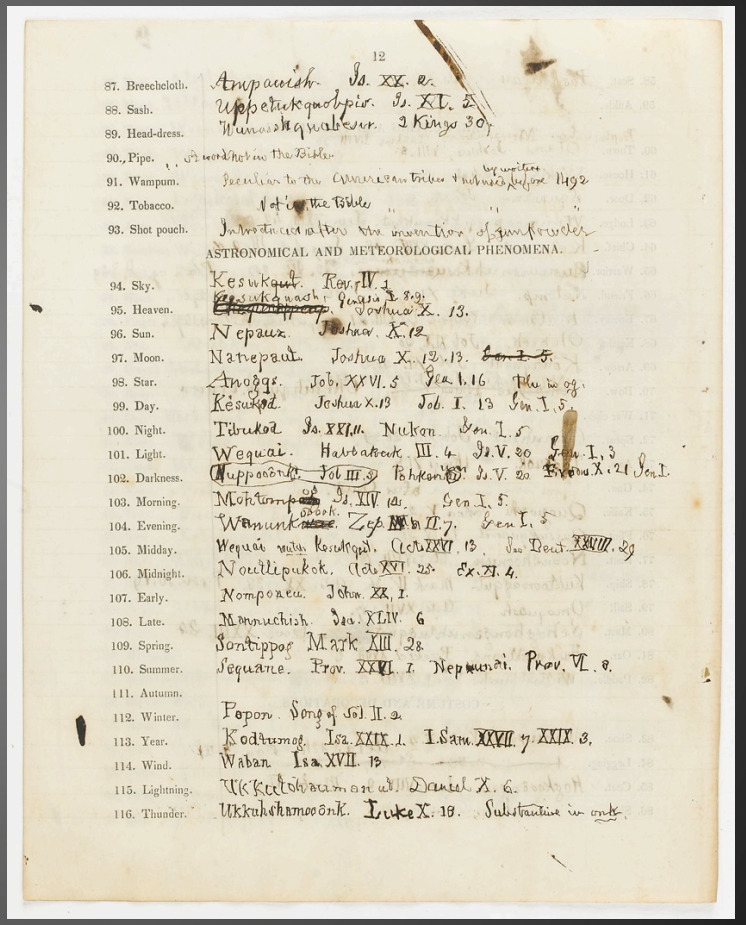
What is Breath of Life

Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators , April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

Smithsonian
National Museum of Natural History

Goals of the Project

- Provide access to digitized materials for participants for their language revitalization efforts
- Create high-resolution digital surrogates for preservation of our collection



12

87. Breechcloth. Ampauish. Is. XX. 2.
88. Sash. Uppetuckquohpir. Is. XL. 5.
89. Head-dress. Wunnashquatteer. 2 Kings 30.
90. Pipe. Not mentioned in the Bible.
91. Wampum. Peculiar to the American tribes & unknown before 1492
92. Tobacco. Not in the Bible.
93. Shot pouch. Introduced after the invention of gunpowder.

ASTRONOMICAL AND METEOROLOGICAL PHENOMENA.

94. Sky. Keswkqut. Rev. IV. 1.
95. Heaven. ~~Keswkquash~~. Joshua X. 13.
96. Sun. Nepaux. Joshua X. 12.
97. Moon. Natnepaut. Joshua X. 12. 13.
98. Star. Anogqs. Job. XXVI. 5 Gen. 1. 16
99. Day. Kesukod. Joshua X. 13 Job. I. 13 Gen. I. 5.
100. Night. Tibukod Is. XXI. 11. Nukon Gen. I. 5
101. Light. Wequai. Habbakuk. III. 4 Is. V. 20 Gen. I. 3
102. Darkness. ~~Nuppaconk~~ Job III. 5 Pohkonk Is. V. 20 Prov. X. 21 Gen. I.
103. Morning. Mohtompog Is. XIV 12. Gen. I. 5.
104. Evening. Wonunk~~ook~~. Zeph. II. 7. Gen. I. 5
105. Midday. Wequai natn keswkqut. Acts XXVI. 13. Deut. XXVIII. 29
106. Midnight. Noellipukok. Acts XVI. 25. Ex. XI. 4.
107. Early. Nompoaeu. John XX. 1.
108. Late. Mannuchish. Isa. XLIV. 6
109. Spring. Sontippog Mark XIII. 28
110. Summer. Sequance. Prov. XXVI. 1. Nepannai. Prov. VI. 8.
111. Autumn.
112. Winter. Popon. Song of Sol. II. 2.
113. Year. Kodtumog. Isa. XXIX. 1. I Sam. XXVII. 7. XXIX. 3.
114. Wind. Waban Isa. XVII. 13
115. Lightning. Ukkuhkuhamoonk. Daniel X. 6.
116. Thunder. Ukkuhkhamoonk. Luke X. 18. Substantive in onk.

NAA MS 1827, "Massachusetts or Natick vocabulary"

Smithsonian
National Museum of Natural History

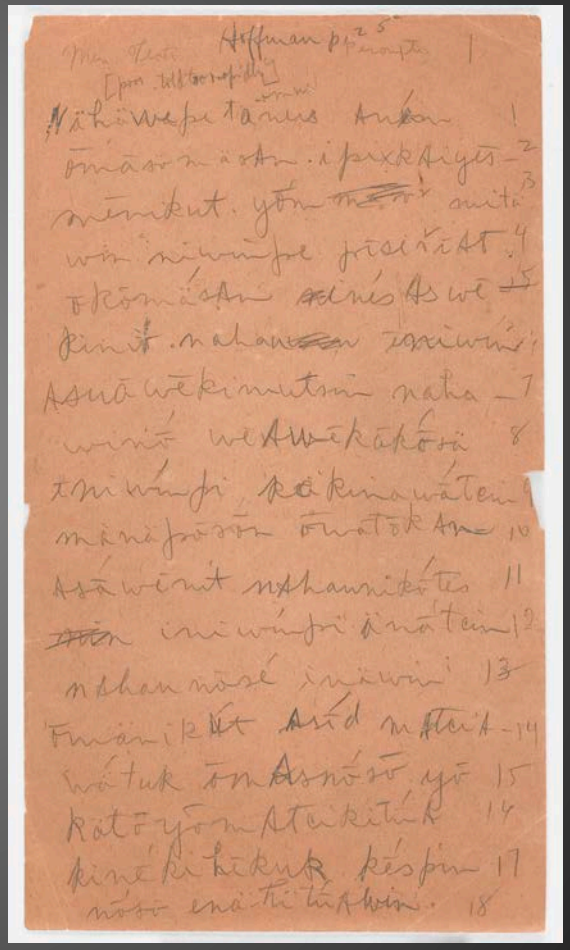
Collaborators

- Digital Lab Staff
  - Digital Program Manager
  - Digital Imaging Technician
  - Digital Imaging Intern
- NAA Staff
  - Reference Staff and Interns
  - Conservator
- Curator
- Program Staff
- Participants

Smithsonian
National Museum of Natural History

Breath of Life 2013 – By the Numbers

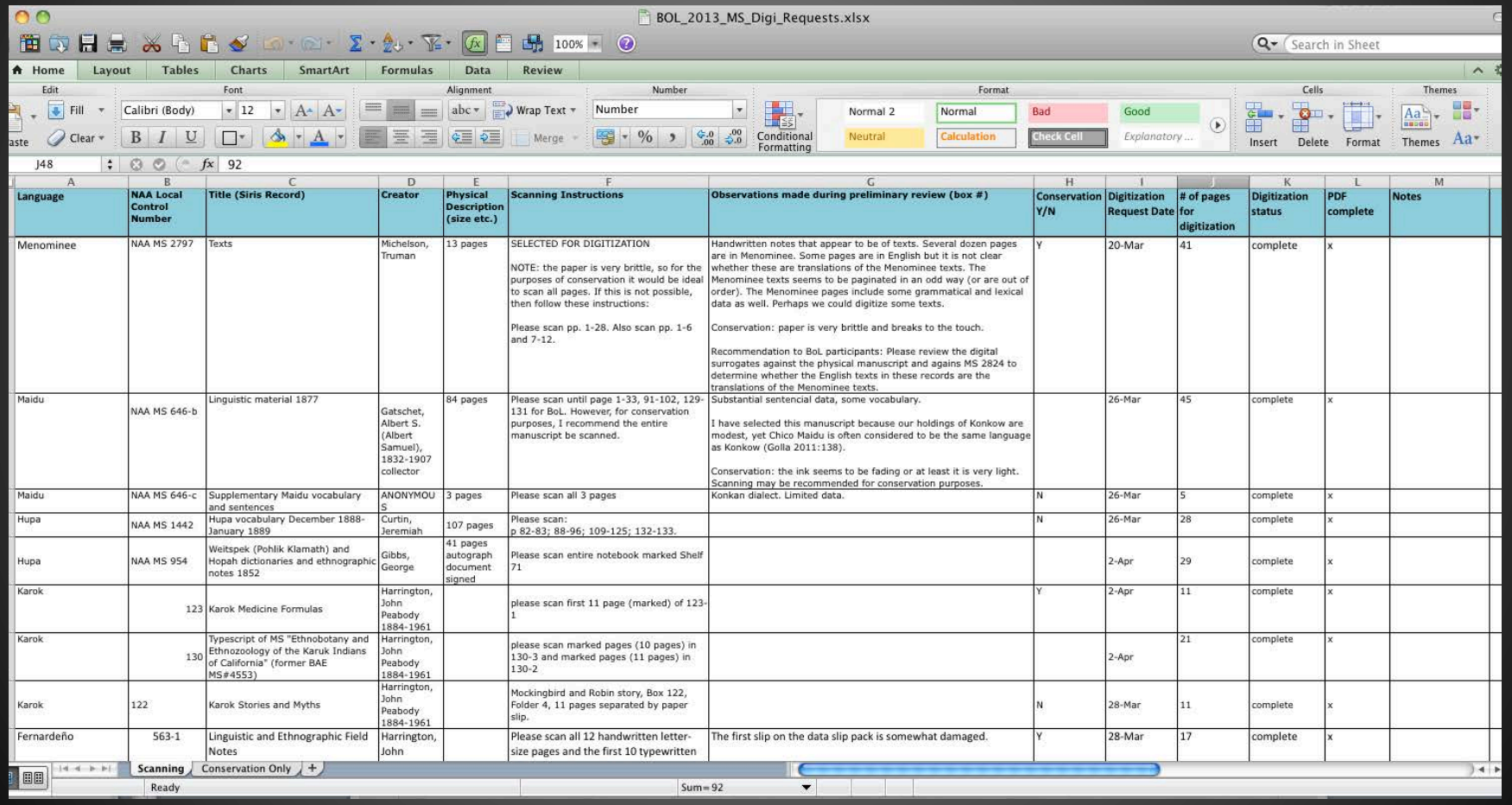
- 16 language teams
- 17 manuscripts, 20 text selections from the John P. Harrington collection, and 7 previously digitized manuscripts
- 1000 pages of archival manuscript material requested; over 3000 digitized in total
- Material ranges from manuscript pages, bound notebooks of various sizes and binding styles, to notecards and slipfiles



NAA MS 2797, "Menominee linguistic notes and texts collected by Truman Michelson, 1910 "

Smithsonian
National Museum of Natural History

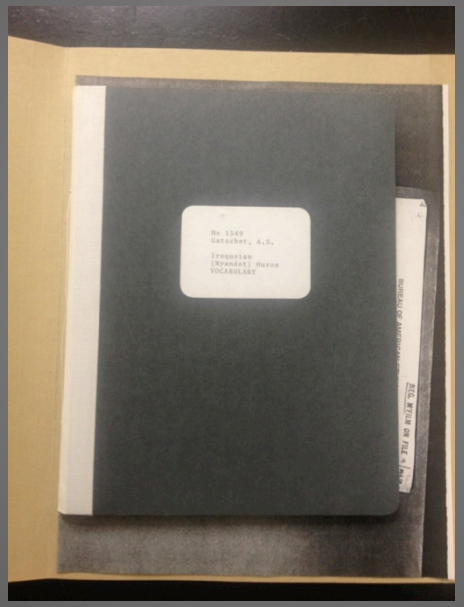
Selection Process

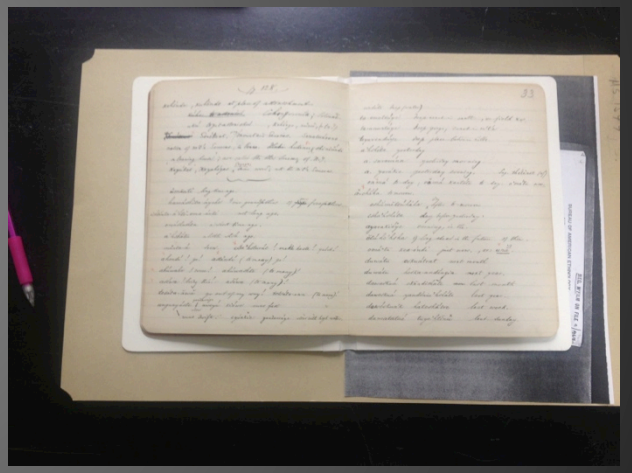

BOL_2013_MS_Digi_Requests.xlsx

Language	NAA Local Control Number	Title (Siris Record)	Creator	Physical Description (size etc.)	Scanning Instructions	Observations made during preliminary review (box #)	Conservation Y/N	Digitization Request Date	# of pages for digitization	Digitization status	PDF complete	Notes
Menominee	NAA MS 2797	Texts	Michelson, Truman	13 pages	SELECTED FOR DIGITIZATION NOTE: the paper is very brittle, so for the purposes of conservation it would be ideal to scan all pages. If this is not possible, then follow these instructions: Please scan pp. 1-28. Also scan pp. 1-6 and 7-12.	Handwritten notes that appear to be of texts. Several dozen pages are in Menominee. Some pages are in English but it is not clear whether these are translations of the Menominee texts. The Menominee texts seems to be paginated in an odd way (or are out of order). The Menominee pages include some grammatical and lexical data as well. Perhaps we could digitize some texts. Conservation: paper is very brittle and breaks to the touch. Recommendation to BoL participants: Please review the digital surrogates against the physical manuscript and agains MS 2824 to determine whether the English texts in these records are the translations of the Menominee texts.	Y	20-Mar	41	complete	x	
Maidu	NAA MS 646-b	Linguistic material 1877	Gatschet, Albert S. (Albert Samuel), 1832-1907 collector	84 pages	Please scan until page 1-33, 91-102, 129-131 for BoL. However, for conservation purposes, I recommend the entire manuscript be scanned.	Substantial sentencial data, some vocabulary. I have selected this manuscript because our holdings of Konkow are modest, yet Chico Maidu is often considered to be the same language as Konkow (Golla 2011:138). Conservation: the ink seems to be fading or at least it is very light. Scanning may be recommended for conservation purposes.		26-Mar	45	complete	x	
Maidu	NAA MS 646-c	Supplementary Maidu vocabulary and sentences	ANONYMOUS	3 pages	Please scan all 3 pages	Konkan dialect. Limited data.	N	26-Mar	5	complete	x	
Hupa	NAA MS 1442	Hupa vocabulary December 1888-January 1889	Curtin, Jeremiah	107 pages	Please scan: p 82-83; 88-96; 109-125; 132-133.		N	26-Mar	28	complete	x	
Hupa	NAA MS 954	Weitspek (Pohlik Klamath) and Hopah dictionaries and ethnographic notes 1852	Gibbs, George	41 pages autograph document signed	Please scan entire notebook marked Shelf 71			2-Apr	29	complete	x	
Karok	123	Karok Medicine Formulas	Harrington, John Peabody 1884-1961		please scan first 11 page (marked) of 123-1		Y	2-Apr	11	complete	x	
Karok	130	Typescript of MS "Ethnobotany and Ethnozoology of the Karuk Indians of California" (former BAE MS#4553)	Harrington, John Peabody 1884-1961		please scan marked pages (10 pages) in 130-3 and marked pages (11 pages) in 130-2			2-Apr	21	complete	x	
Karok	122	Karok Stories and Myths	Harrington, John Peabody 1884-1961		Mockingbird and Robin story, Box 122, Folder 4, 11 pages separated by paper slip.		N	28-Mar	11	complete	x	
Fernardeño	563-1	Linguistic and Ethnographic Field Notes	Harrington, John		Please scan all 12 handwritten letter-size pages and the first 10 typewritten	The first slip on the data slip pack is somewhat damaged.	Y	28-Mar	17	complete	x	

Smithsonian
National Museum of Natural History

Breath of life 2013 – variations on a theme



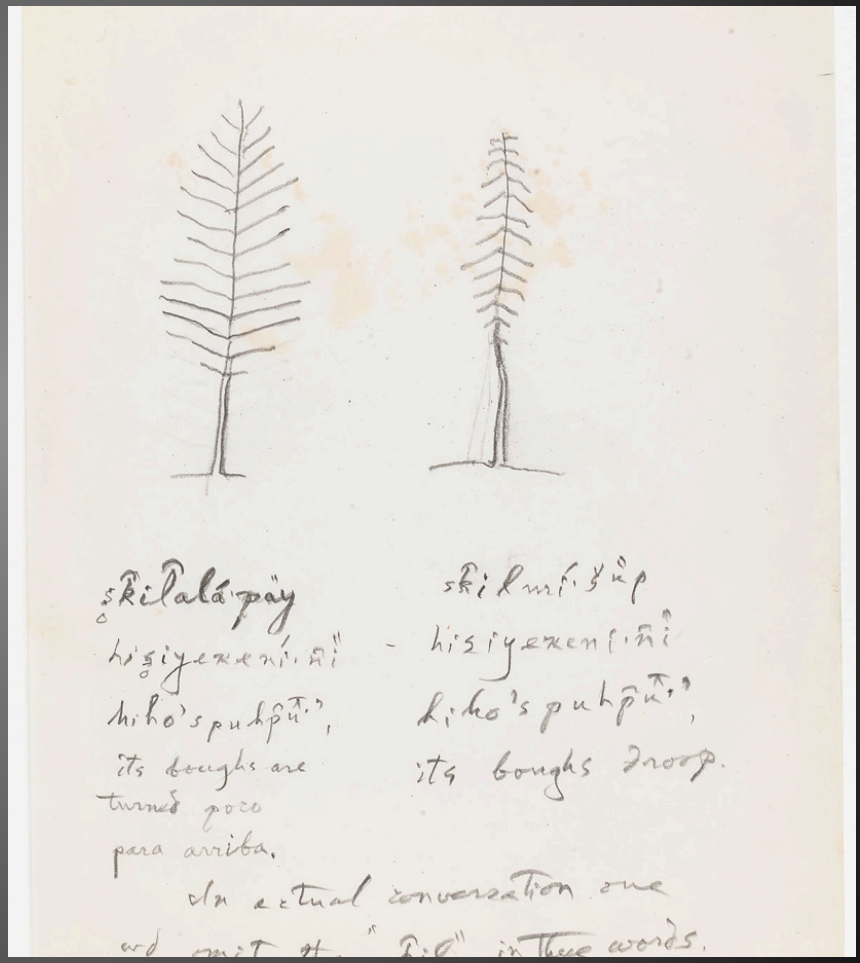


NAA MS 1549, "Wandot/Wyandot/ vocabulary January 8, February 7, March 15, 1881"

Smithsonian
National Museum of Natural History

The Harrington Issue

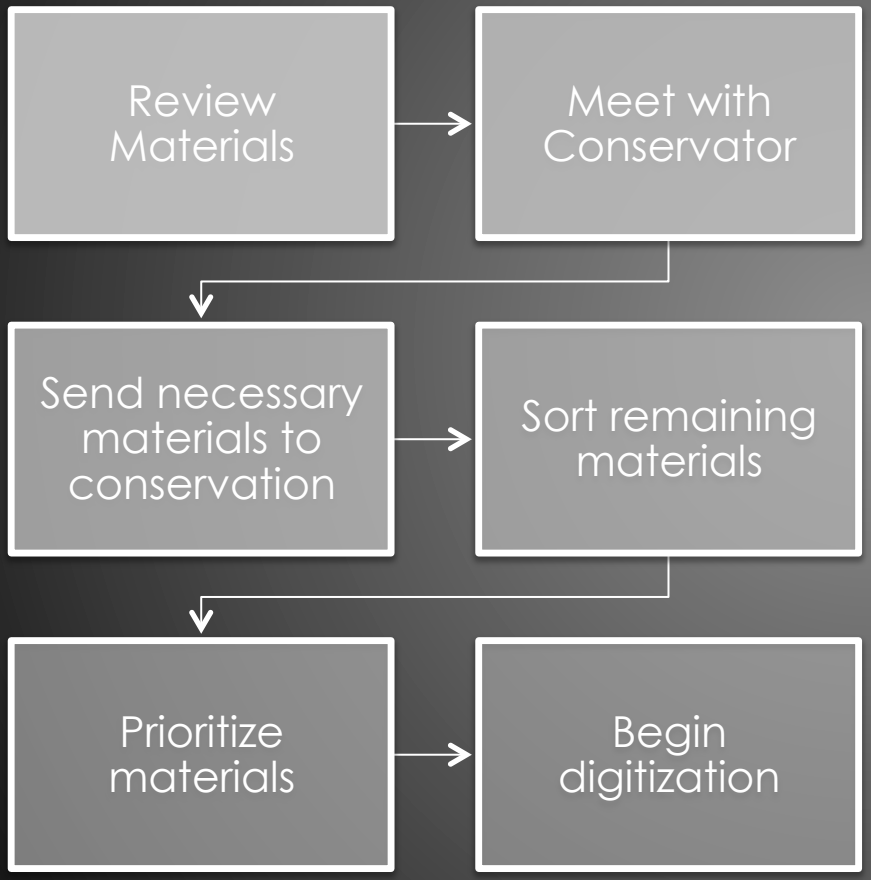
- Extremely large, not fully processed collection
- In the middle of another digitization project of the collection
- Approach differently than other materials
  - Not digitizing full folders or boxes for preservation purposes (access only)

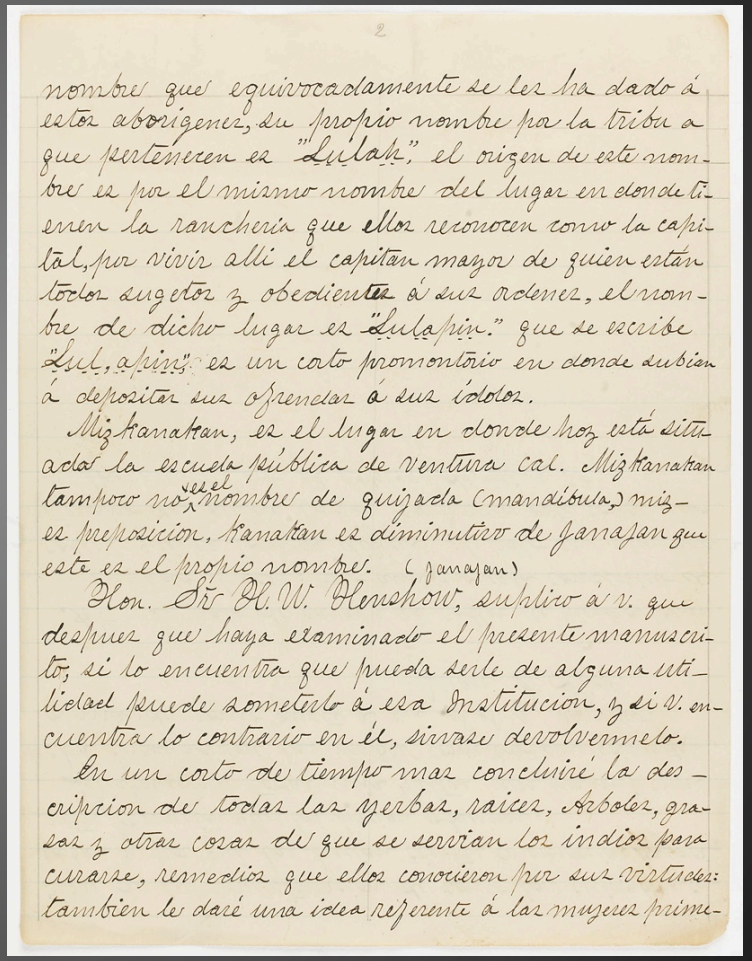


John P. Harrington collection Box 483, Folder 1

Smithsonian
National Museum of Natural History

Prioritization





nombre que equivocadamente se les ha dado á estos aborigenes, su propio nombre por la tribu a que pertenecen es "Lulak", el origen de este nombre es por el mismo nombre del lugar en donde tienen la ranchería que ellos reconocen como la capital, por vivir allí el capitan mayor de quien están todos sugetos y obedientes á sus ordenes, el nombre de dicho lugar es "Lulapin." que se escribe "Lul, apin" es un corto promontorio en donde suben á depositar sus ofrendas á sus ídolos.

Miskanakan, es el lugar en donde hoy está situada la escuela pública de Ventura Cal. Miskanakan tampoco no es el nombre de quijada (mandíbula,) miz- es preposicion, kanakan es diminutivo de Janajan que este es el propio nombre. (janajan)

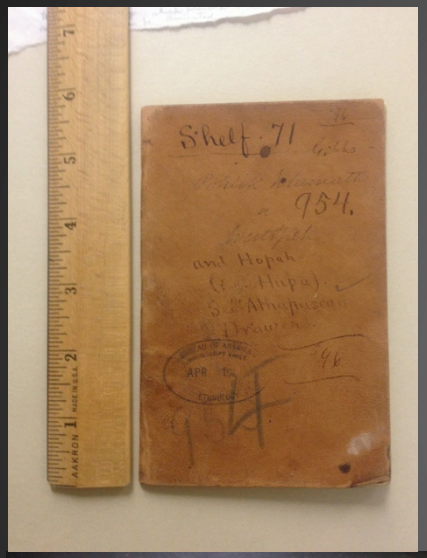
Hon. Sr. H. W. Henshow, suplico á v. que despues que haya examinado el presente manuscrito, si lo encuentra que pueda serle de alguna utilidad puede someterlo á esa Institucion, y si v. encuentra lo contrario en él, sírvase devolvermelo.

En un corto de tiempo mas concluiré la descripcion de todas las yerbas, raices, Arboles, grasas y otras cosas de que se servian los indios para curarse, remedios que ellos conocieron por sus virtudes: tambien le daré una idea referente á las mujeres prime-

NAA MS 3718, "Fourth of July, 1890 "

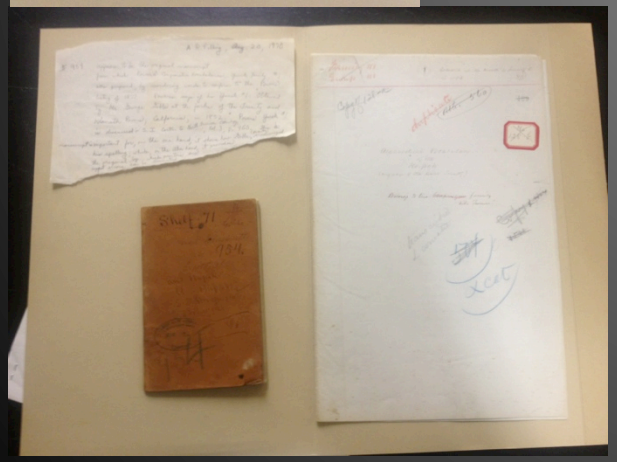
Smithsonian
National Museum of Natural History

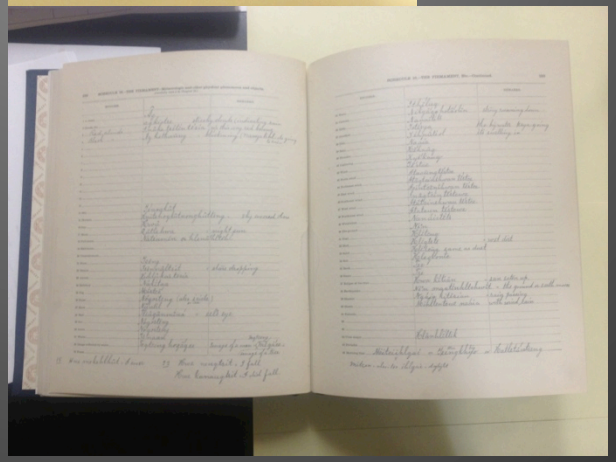
Materials: What we're working with

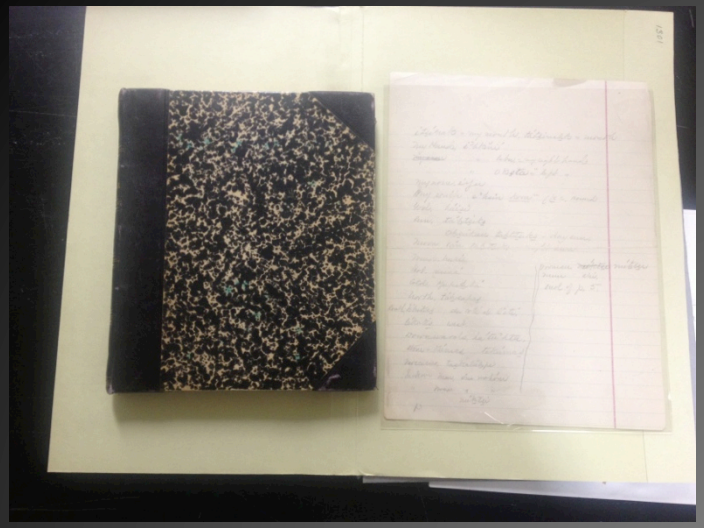


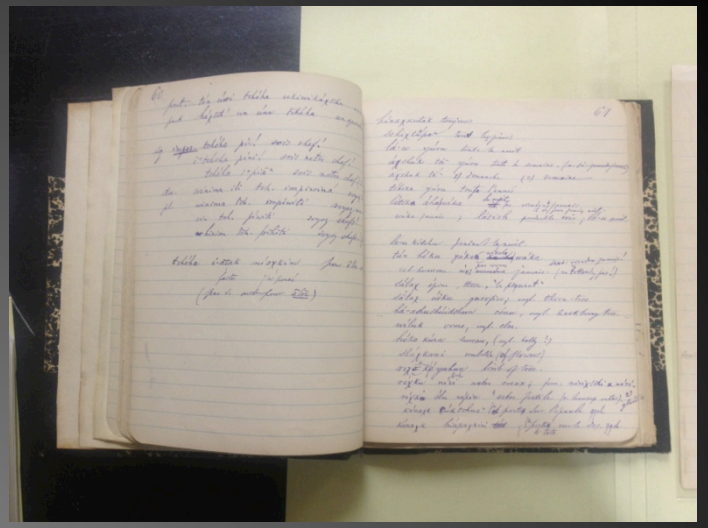
NAA MS 954

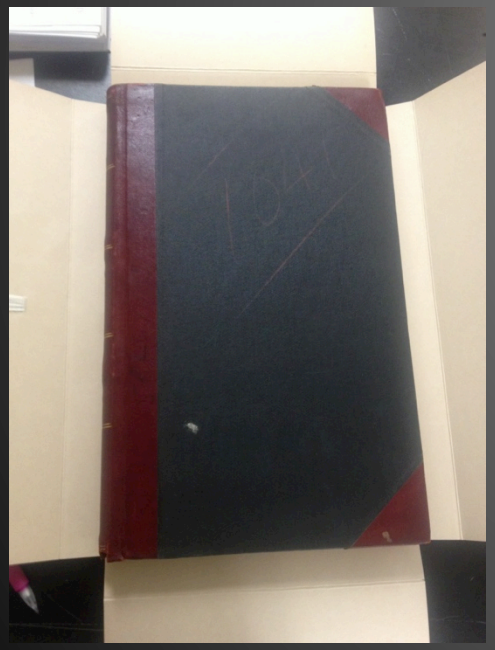
NAA MS 1442

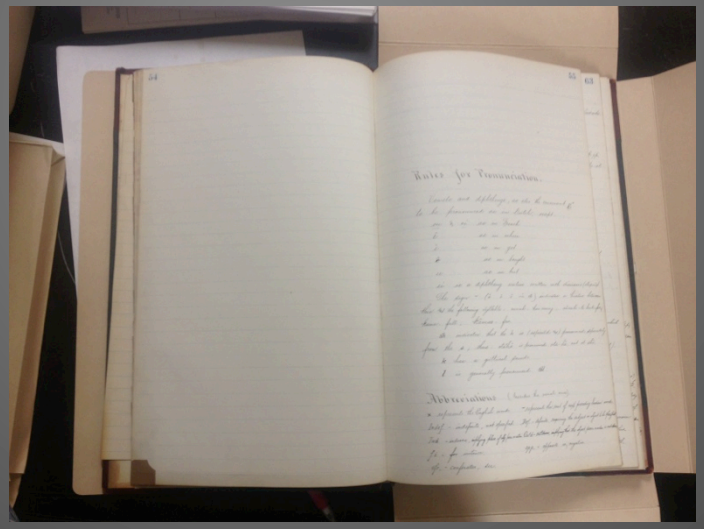


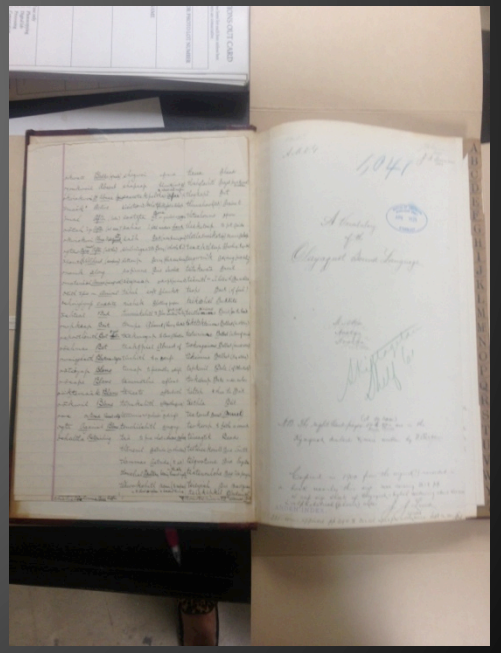








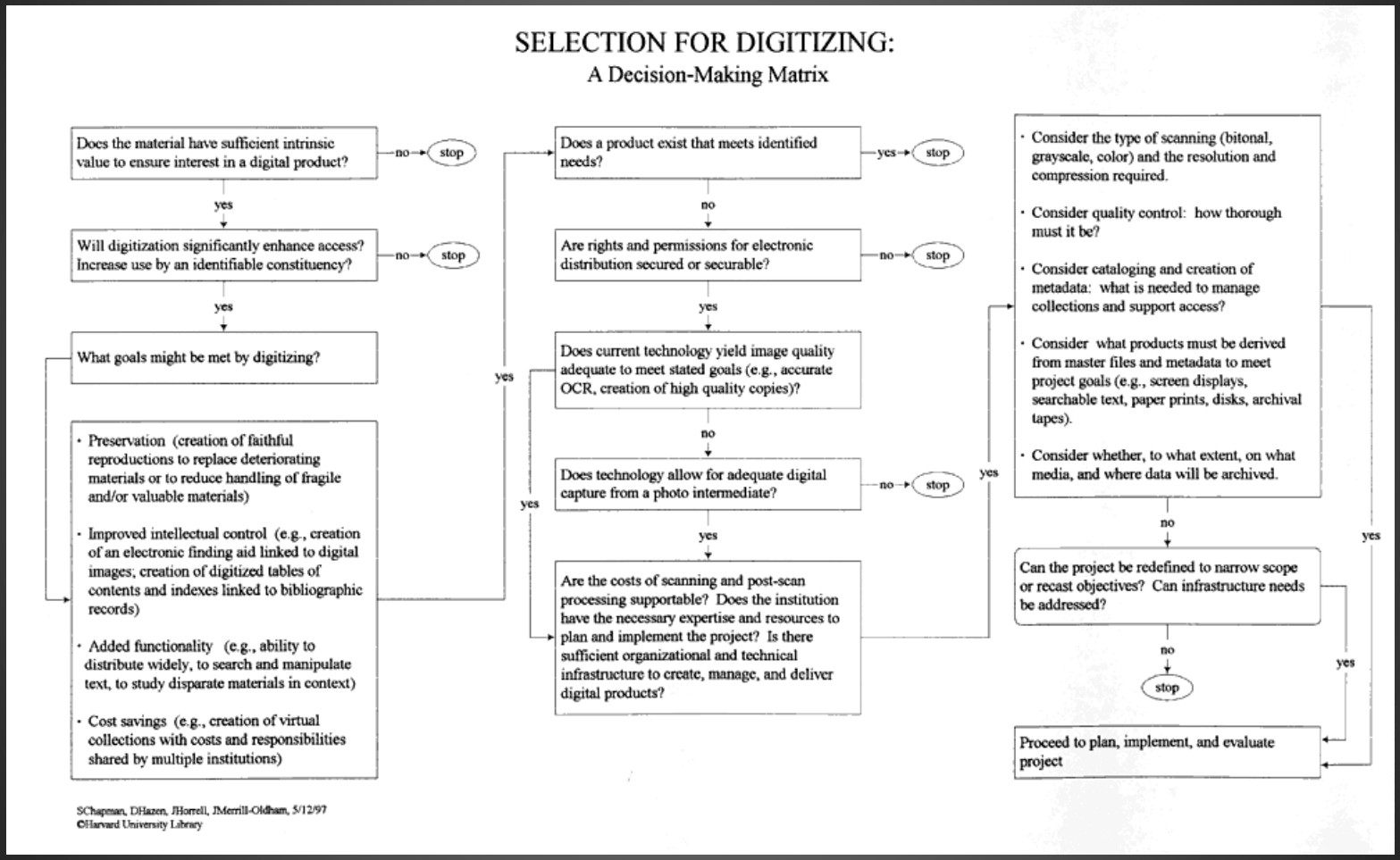




Above image from: NAA MS 1301 (T) and NAA MS 1041 (B)

Smithsonian
National Museum of Natural History

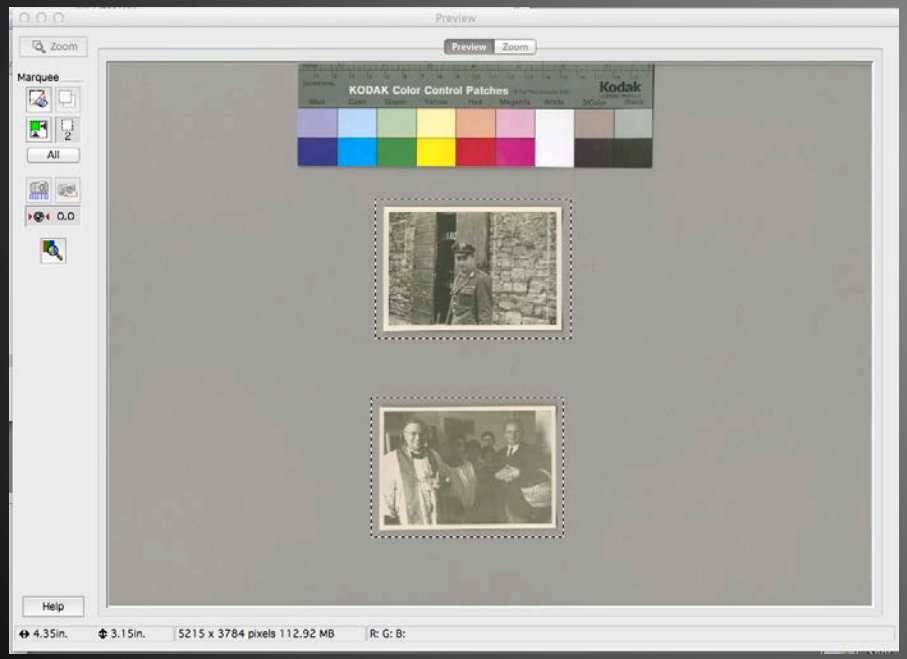
A textbook approach



Hzen, Dan, Horrell, Jeffery, and Merrill-Oldham. (1998) "Selection for Digitizing: A Decision-Making Matrix," Selecting Research Collections for Digitization – Full Report.

Smithsonian
National Museum of Natural History

Image Creation: Recommendations and Realities



Smithsonian
National Museum of Natural History

Questions to Consider

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

Smithsonian
National Museum of Natural History

Federal Agencies Digitization Guidelines Initiative

Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators ,
April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

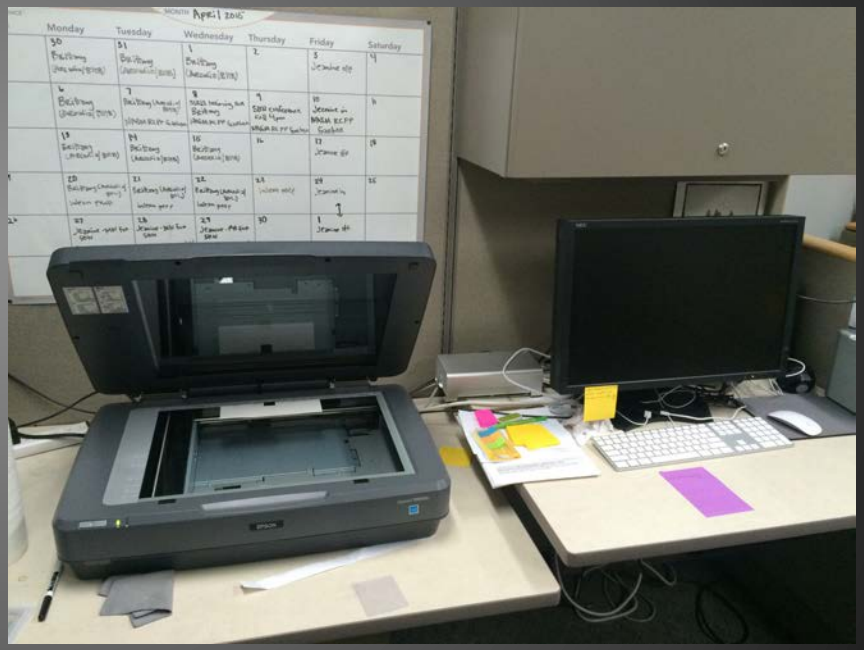
(Even more!) things to consider

- Staff and Time
  - Expertise and Testing
- Imaging environment
  - Space
  - Lights
  - Monitor
- Hardware and Software
  - Scanner v. Camera
  - Platform (Mac v. PC)
  - Imaging Software
    - Proprietary with equipment
    - Adobe products
    - Image Science Associates

BOTTOM LINE: $$$$$$$$$$$$$$$$$$

Smithsonian
National Museum of Natural History

Scanner v. Camera



Smithsonian
National Museum of Natural History

Technical Specifications

- Bit depth
- Tone/Endpoints
- Color mode/profile
- Resolution
- File formats

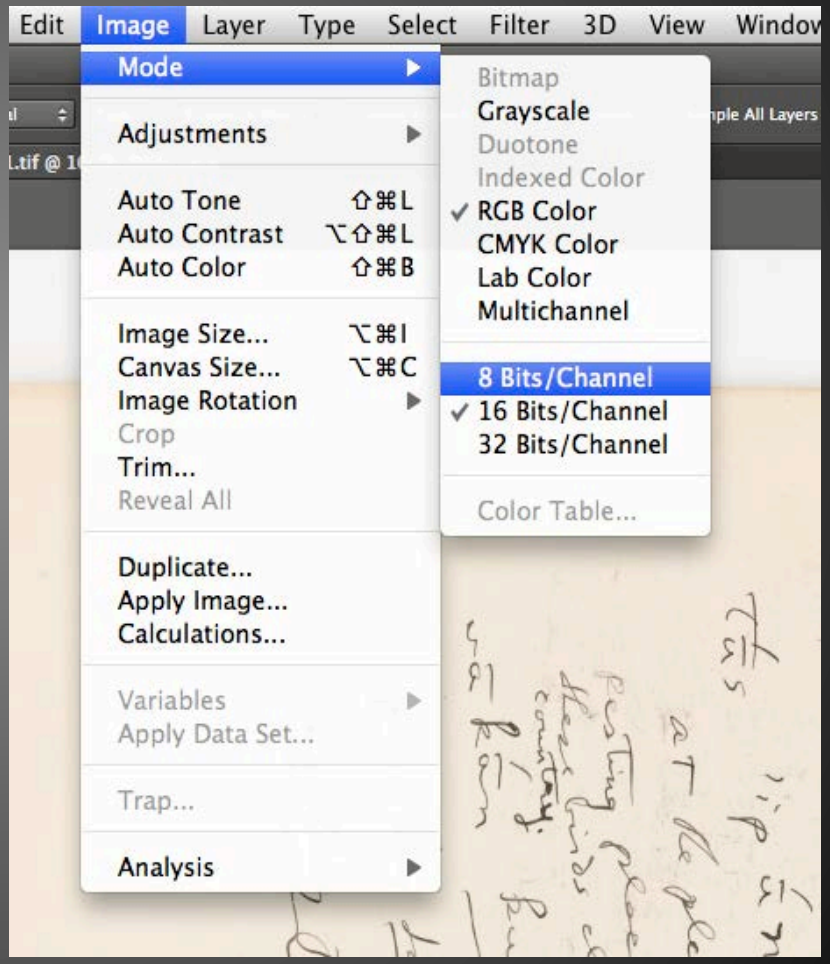
Technical Specifications – Bit Depth

- Bit depth: the color information stored in an image; the higher the bit depth, the more colors an image can store. A 1 bit image only shows two colors – black and white
  - 8 bit: 2^8= 256 colors
  - 16 bit: 2^16 = 65,536 colors
  - 24 bit: 2^24 = 16,000,000+ colors
  - 48 bit: 2^48 = 3,000,000,000 colors

Smithsonian
National Museum of Natural History

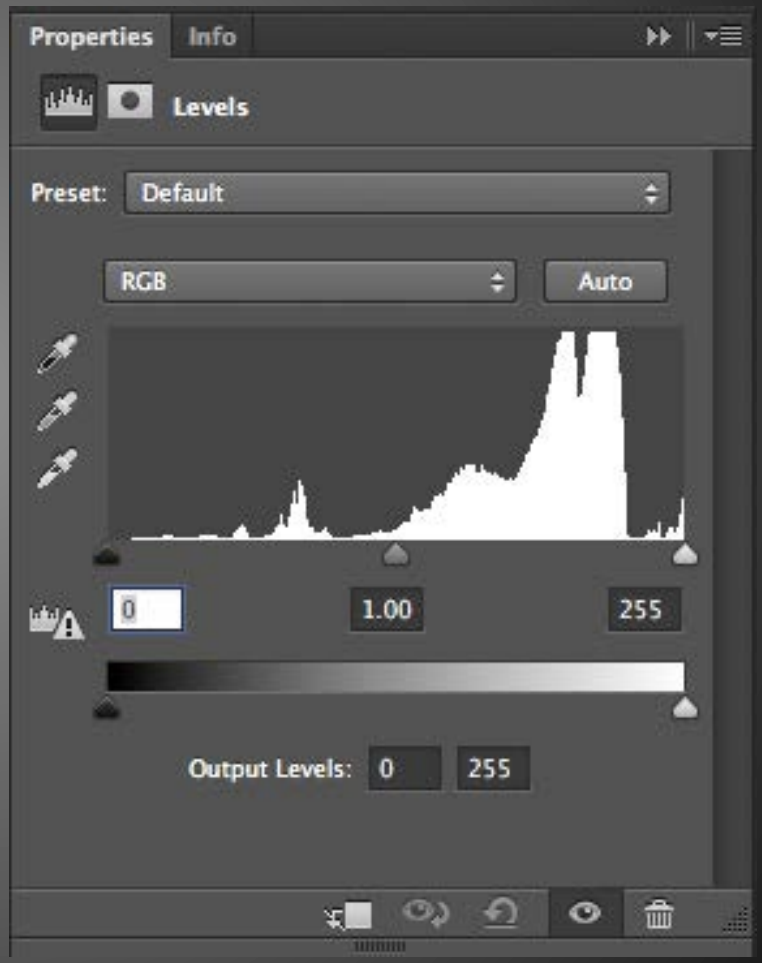
Bit Depth – Recommendations

- Capture images in a high bit depth (16 bit or higher) then convert to a lower bit depth as needed
- Higher bit depth images will increase accuracy of representing the original BUT will create larger files



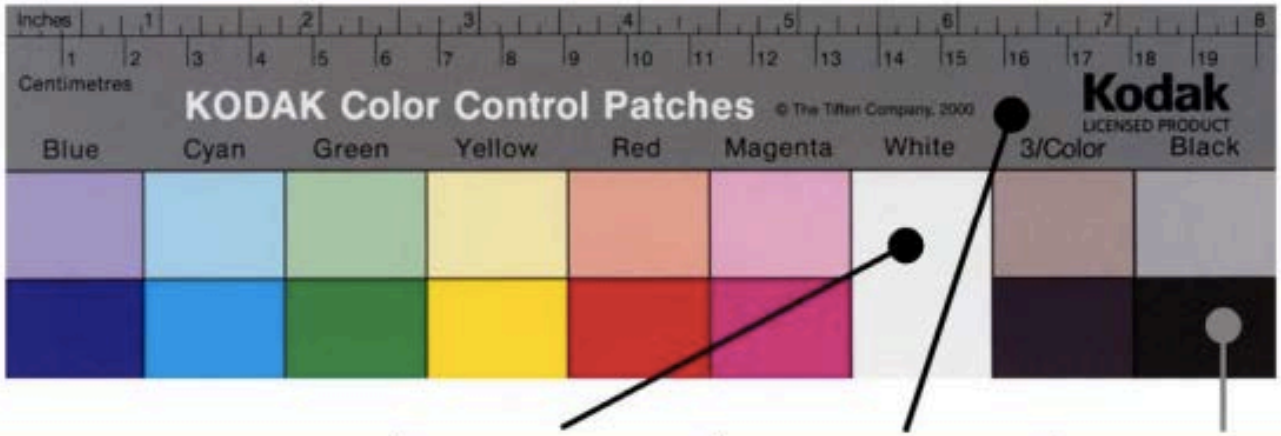
Technical Specifications – Tone

- Each pixel in an image file is assigned a value between 0 and 255, with 0 being a "true" blackest black, and 255 being a "true" whitest white.
- Values below 9 (too black) or above 247 (too white) lose all detail



Smithsonian
National Museum of Natural History

Tone - Recommendations



		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%

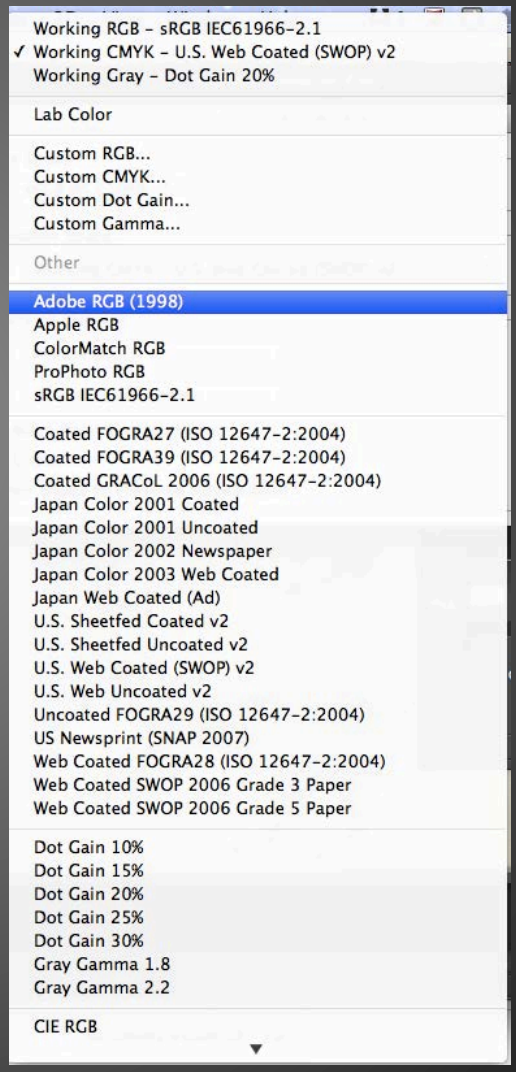
*Aimpoint for mid point (MP) to be calculated from actual values for white point (WP) and black point (BP) using the following formula: MP = WP – 0.63(WP – BP)

Technical Specifications – Color Mode

- "Grayscale image files consist of a single channel, commonly either 8-bits (256 levels) or 16-bits (65,536 levels) per pixel with the tonal values ranging from black to white. Color images consist of three or more grayscale channels that represent color and brightness information. Common color modes include RGB (red, green, blue), CMYK (cyan, magenta, yellow, black), and LAB (lightness, red-green, blue-yellow). The channels in color files may be either 8- bits (256 levels) or 16-bits (65,536 levels). Display and output devices mathematically combine the numeric values from the multiple channels to form full color pixels, ranging from black to white and to full colors."

Color Mode – Recommendations

- Adobe RGB 1998 (or Adobe sRGB, alternatively) is the recommended color profile for master files



Technical Specifications – Resolution

- Resolution: the number of pixels in each dimension that can be displayed
- DPI: dots per inch (more appropriate for printing)
- PPI: pixels per inch (more accurately describes images)
- "Resolution Threshold": the point where no matter how many more pixels added per inch, no more information from original is gained

Smithsonian
National Museum of Natural History

Resolution – Recommendations

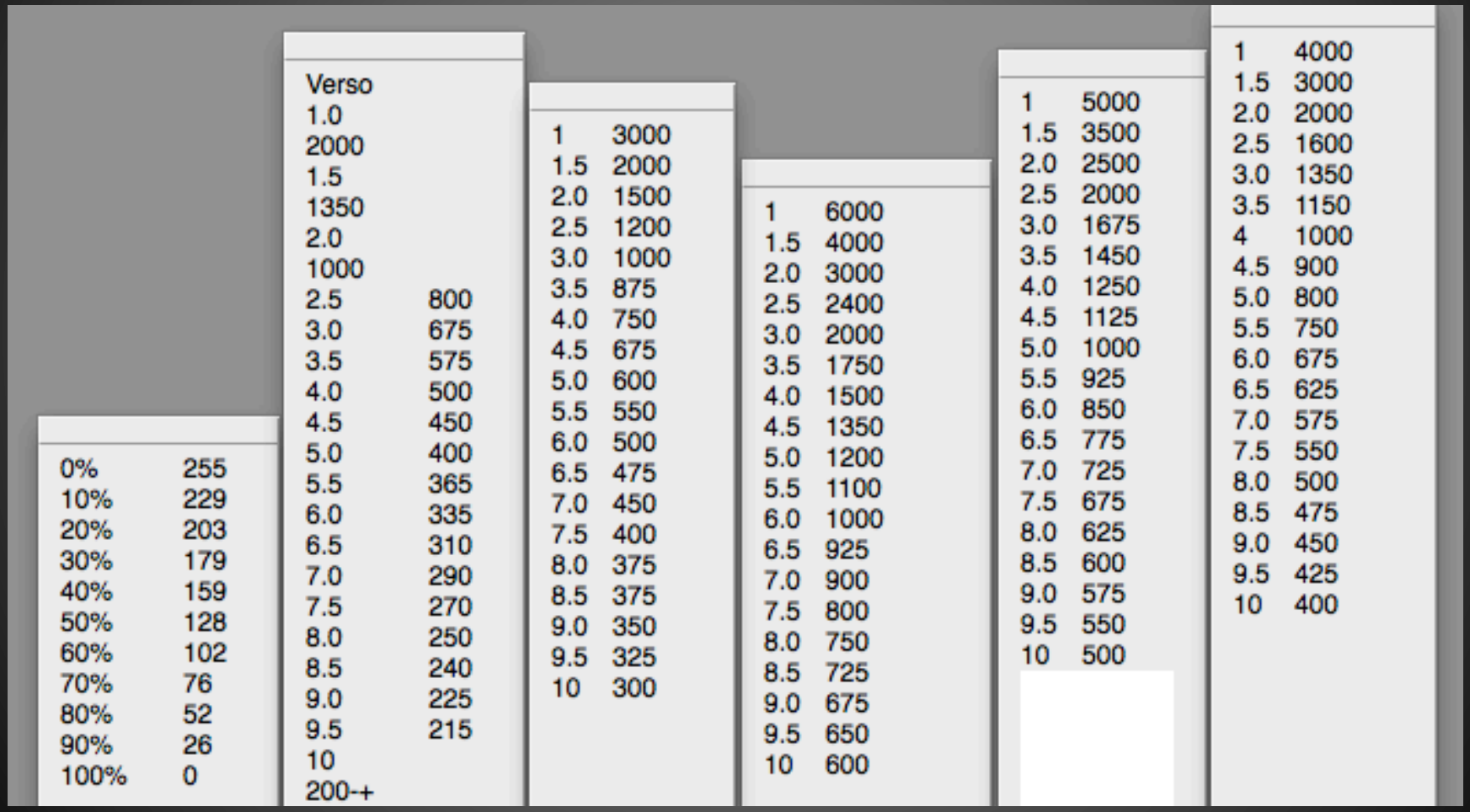
"All digital images created from reflective analog textual materials should have a minimum of 4,000 pixels along their longest dimension, and a minimum resolution of 400 dpi. Text on the reverse side of an object may be scanned at 200 dpi. All reflective photographic materials are at a minimum of 5,000
pixels on their longest dimension.

All digital images created from transmissive analog materials should have a minimum of 6,000 pixels along their longest dimension, and a minimum of 600 dpi.

For example, an 8 x 10-inch photograph should be scanned at 500 dpi; a 5-inch photograph at 1000 dpi, etc. However, photographic items should never be scanned at less than 500 dpi — even if the longest dimension of the object is greater than 10 inches. When measuring the longest dimension of an image, round down to the nearest half inch. When calculating resolution, round up to the closest multiple of 25. Calculations should reflect the dimensions of actual item to be scanned, not its matting or support."

Smithsonian
National Museum of Natural History

Resolution – Recommendations



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

Smithsonian
National Museum of Natural History

Technical Specifications – File Formats

File Format	Technical Considerations	Recommended Use
TIFF	▪ "De facto" raster image format used for master files ▪ Simply encoded raster-based format ▪ Accommodates internal technical metadata in header/extensible and customizable header tags ▪ Supports Adobe's XMP (Extensible Metadata Platform) ▪ Accommodates large number of color spaces and profiles ▪ Supports device independent color space (CIE L*a*b) ▪ Uncompressed; lossless compression (Supports multiple compression types for 1-bit files). JPEG compression not recommended in TIFF file ▪ High-bit compatible ▪ Can support layers, alpha channels ▪ Accommodates large file sizes ▪ Anticipate greater preservation support in repository settings; preferred raster image format for preservation ▪ Widely supported and used ▪ Long track record (format is over 10 years old) ▪ Potential loss of Adobe support of TIFF in favor of PDF? ▪ Not suitable as access file–no native support in current web browsers	Preferred format for production master file
JPEG 2000	▪ Increasingly considered as a viable format for master image files, but not yet widely adopted ▪ More complex model for encoding data (content is not saved as raster data) ▪ Supports multiple resolutions ▪ Extended version supports color profiles ▪ Extended version supports layers ▪ Includes additional compression algorithms to JPEG (wavelet, lossless)	Rapidly gaining acceptance as a format for production master file–however, not currently widely implemented (or TIFF files are also being saved alongside JPEG 2000 files as master formats (?))
JFIF/JPEG	▪ Lossy compression, but most software allows for adjustable level of compression ▪ Presence of compression artifacts ▪ Smaller files ▪ High-bit compatible ▪ Longer decompression time ▪ Supports only a limited set of internal technical metadata ▪ Supports a limited number of color spaces ▪ Not suitable format for editing image files–saving, processing, and resaving results in degradation of image quality after about 3 saves	Access derivative file use only–not recommended for text or line drawings
PDF	▪ Intended to be a highly structured page description language that can contain embedded objects, such as raster images, in their respective formats. ▪ Works better as a container for multiple logical objects that make up a coherent whole or composite document ▪ More complex format due to embedded/externally linked objects ▪ Implements Adobe's XMP specification for embedding metadata in XML ▪ Can use different compression on different parts of the file; supports multiple compression schemes ▪ Supports a limited number of color spaces	Not recommended for production master files
PDF/A	▪	
PNG	▪ Simple raster format ▪ High-bit compatible ▪ Lossless compression ▪ Supports alpha channels ▪ Not widely adopted by imaging community ▪ Native support available in later web browsers as access file	Possible format for production master file–not currently widely implemented
GIF	▪ Lossy (high color) and lossless compression ▪ Limited color palette ▪ 8-bit maximum, color images are dithered ▪ Short decompression time	Access derivative file use only–recommend for text records
[ASCII]	▪ For image files converted to text ▪ Potential loss to look and feel of document/formatting	N/A
[XML]	▪ For image files converted to text ▪ Hierarchical structure ▪ Good for encoding digital library-like objects or records ▪ Allows for fast and efficient end-user searching for text retrieval ▪ Easily exchanged across platforms/systems	N/A

File Format - Recommendations

File Format	Technical Considerations	Recommended Use
TIFF	▪ "De facto" raster image format used for master files ▪ Simply encoded raster-based format ▪ Accommodates internal technical metadata in header/extensible and customizable header tags ▪ Supports Adobe's XMP (Extensible Metadata Platform) ▪ Accommodates large number of color spaces and profiles ▪ Supports device independent color space (CIE L*a*b) ▪ Uncompressed; lossless compression (Supports multiple compression types for 1-bit files). JPEG compression not recommended in TIFF file ▪ High-bit compatible ▪ Can support layers, alpha channels ▪ Accommodates large file sizes ▪ Anticipate greater preservation support in repository settings; preferred raster image format for preservation ▪ Widely supported and used ▪ Long track record (format is over 10 years old) ▪ Potential loss of Adobe support of TIFF in favor of PDF? ▪ Not suitable as access file–no native support in current web browsers	Preferred format for production master file

Smithsonian
National Museum of Natural History

Basic Scanning Workflow

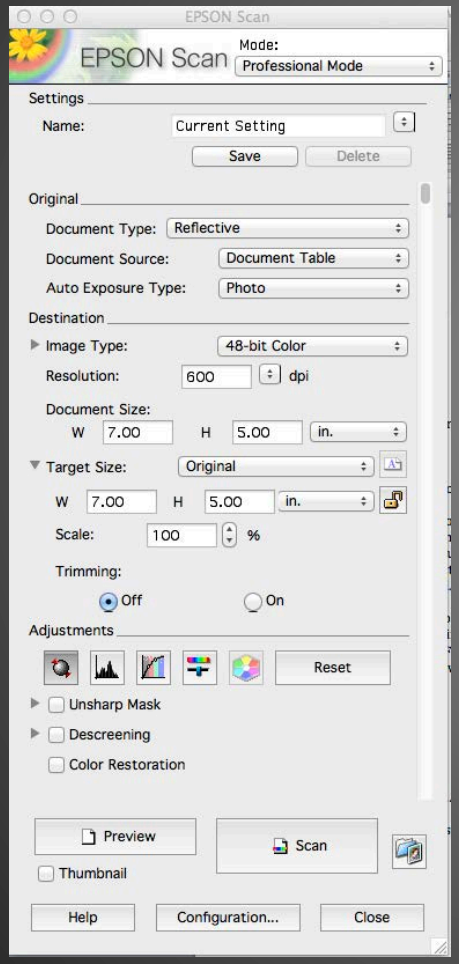
1. Setup Scanning Workstation
2. Select Scanner Settings
3. Setup Scanner for Use
4. Scan
5. Save

Extra: Batch Scanning

1. Setup Scanning Workstation

- Turn on computer and scanner
- Remove any loose articles of clothing or jewelry
- Wash hands or wear gloves (no lotions, no hand sanitizers)
- You will need:
  - A ruler
  - Staging area for original objects
  - Gloves and/or clean hands
  - Neutral background board

2. Select Scanner Settings



- Transmissive (negatives) v. Reflective (photographs, text)
- Resolution
- Color Profile
- Tonal Range
- Turn off adjustments!

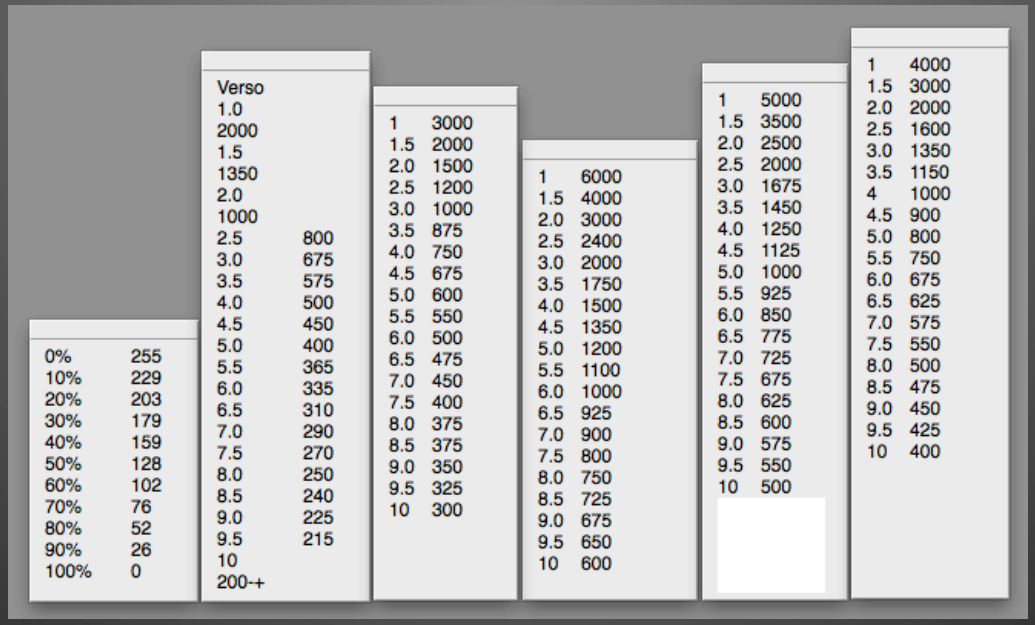
3. Setting up Scanner for Use

- Wipe down scanner glass with lint free cloth
- Place color bar target in appropriate place (center top, for instance)
  - Utilizing a target is crucial for maintaining image quality and to serve as reference point if the integrity of images is every questioned

Smithsonian
National Museum of Natural History

3. Setting up Scanner for Use

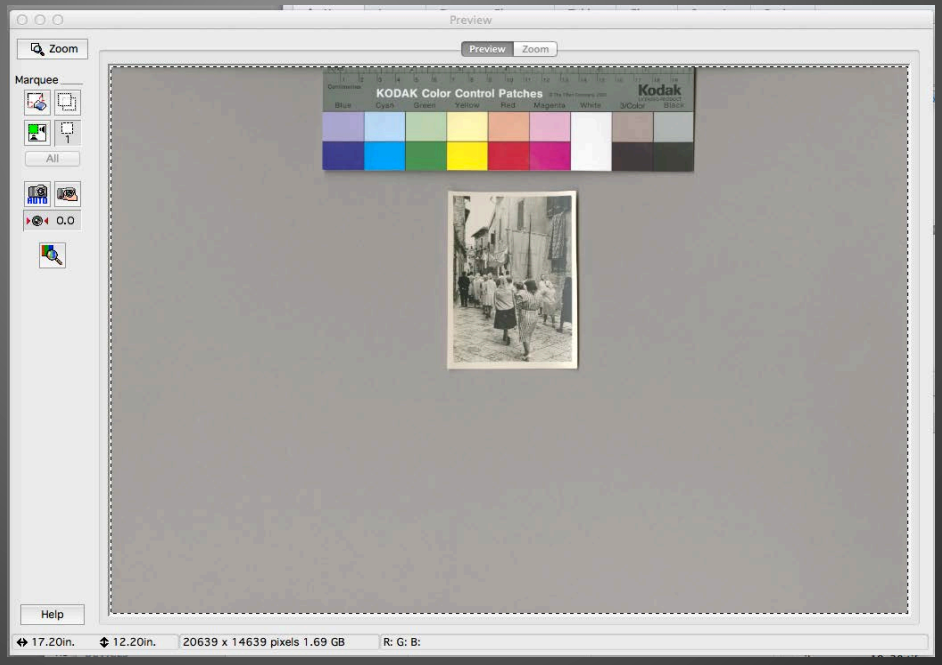
Measure the object you wish to scan. All digital images created from reflective analog textual materials should have a minimum of 4,000 pixels along their longest dimension, When measuring the longest dimension of an image, round down to the nearest half-inch. When calculating resolution, round up to the closest multiple of 25. Calculations should reflect the dimensions of actual item to be scanned, not its matting or support.



Smithsonian
National Museum of Natural History

4. Scan

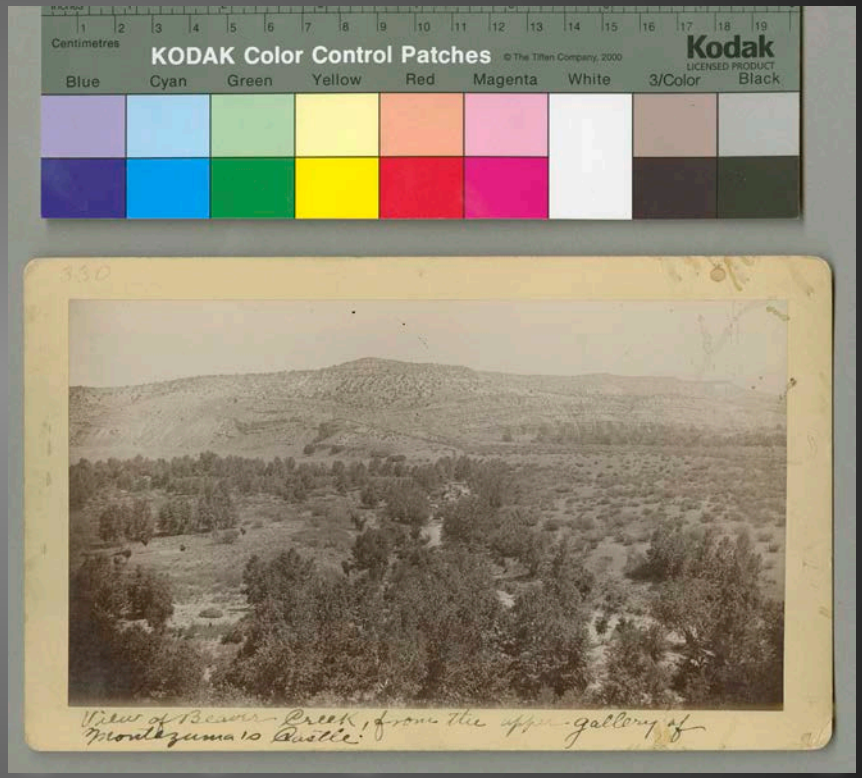
- Once you have determined the correct resolution, place the object on the glass, (for photographic material, face down) about an inch below, in the center of, the color bar. Place the background board over the scanner glass and carefully close the scanner lid.



Smithsonian
National Museum of Natural History

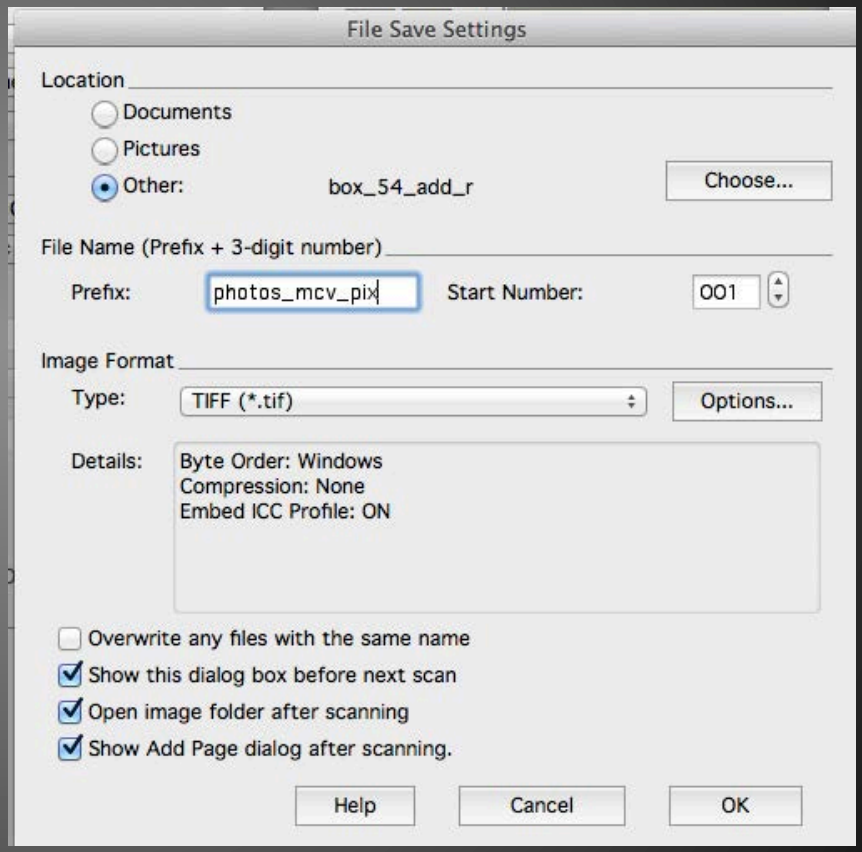
4. Scan

- Check your endpoints (white and black) to make sure you are within the correct range.
- If your endpoints are within the range, select the area to scan – you do not need to scan the entire scanner bed, just the object and the color bar target. Create a box with the dotted line around the color bar target and the object, with about ¼ inch border of grey background around the edge.



5. Save

- Assign appropriate filename and save images into the appropriate folder location



Smithsonian
National Museum of Natural History

Extra – Batch Scanning

- If multiple objects need to be scanned, and more than one fit on the scanner at the same time, they can usually be batched scanned.
- **When determining resolution, set your resolution to align with the SMALLEST OBJECT ON THE GLASS, so that all objects reach the required pixel dimensions.**
- Arrange the objects on the glass far enough apart so you will be able to get a ¼ inch border around each.

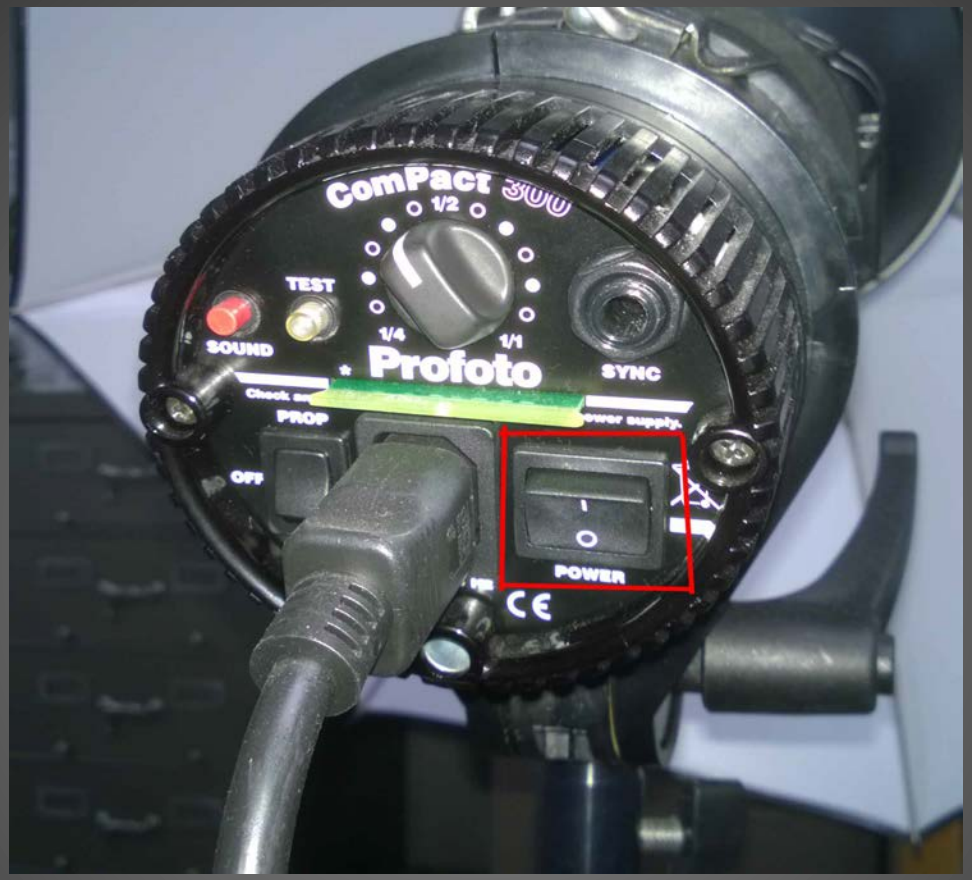
Extra – Batch Scanning

- In the preview window, crop ALL images so there is a dotted border around each.
- When batch scanning, you do not need to include the color bar target in the scans.

Smithsonian
National Museum of Natural History

Bonus: Abridged Camera Workflow

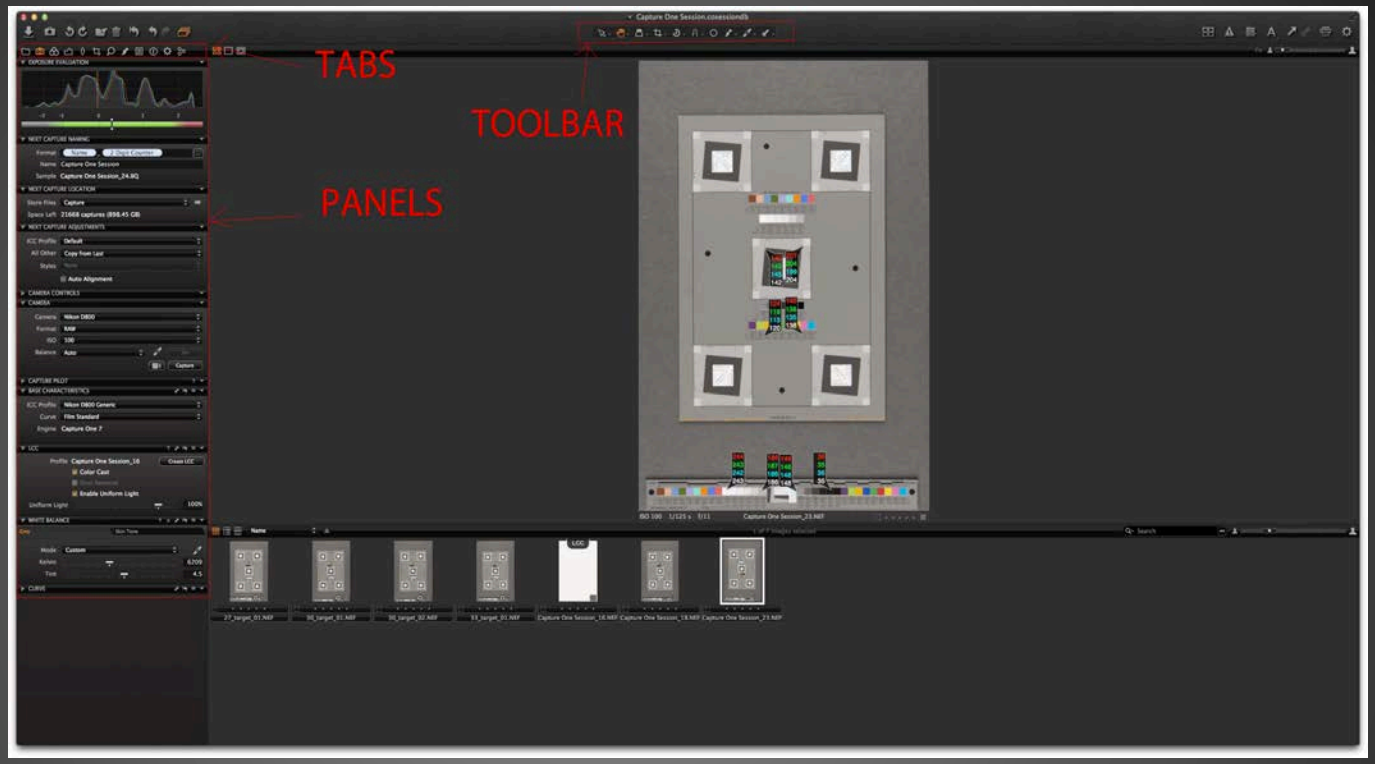
Turn on Nikon camera (slide control on the camera).



Bonus: Abridged Camera Workflow

Turn on both Profoto lighting fixtures (black switches on the base near power cord)

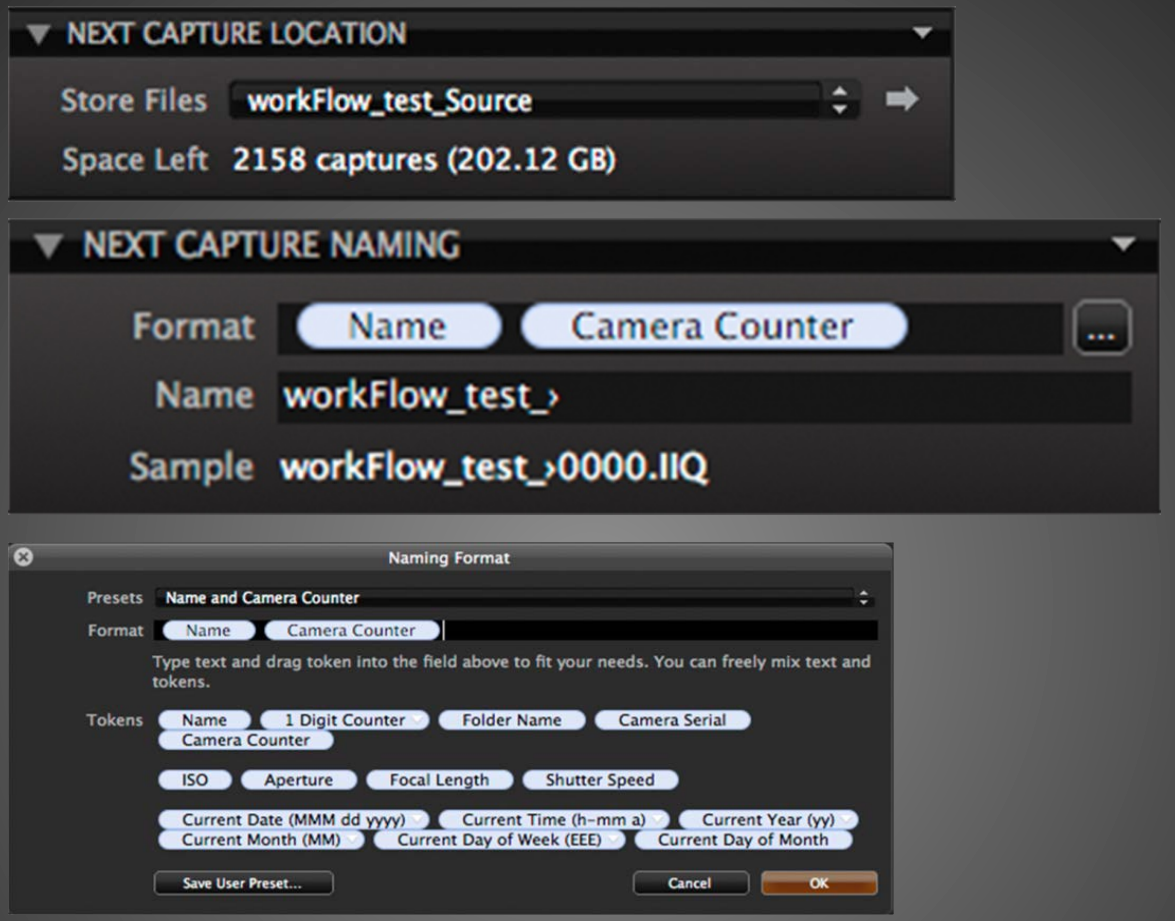
Smithsonian
National Museum of Natural History



Bonus: Abridged Camera Workflow

Launch Capture One software

Smithsonian
National Museum of Natural History



Bonus: Abridged Camera Workflow

Setup camera counter and filenaming in Capture One

Smithsonian
National Museum of Natural History

Bonus: Abridged Camera Workflow

Rotate the 'Focus Ring' on the Nikon's lens to focus. Note that the image will be very pixelated even when it is focused.

Once image is focused, hit command+K or Capture key

Smithsonian
National Museum of Natural History

Bonus: Abridged Camera Workflow

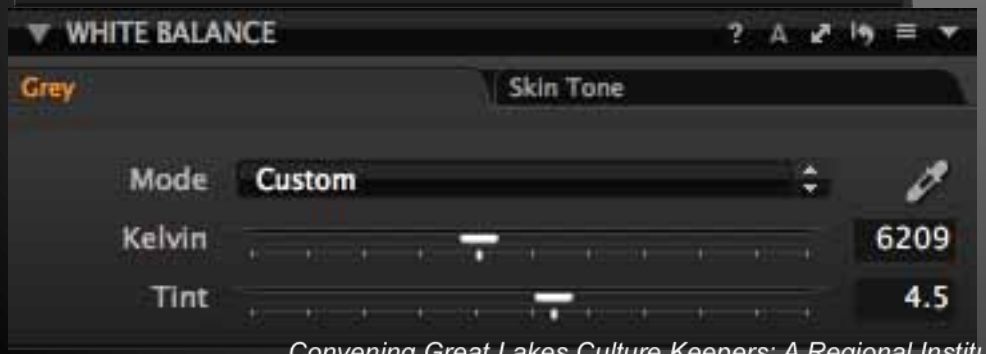
Using the Grayscale on the Device/Object Target, make sure your whites are between 242 and 248 or whatever is appropriate for you subject matter.

Smithsonian
National Museum of Natural History

Create Lens Cast Calibration

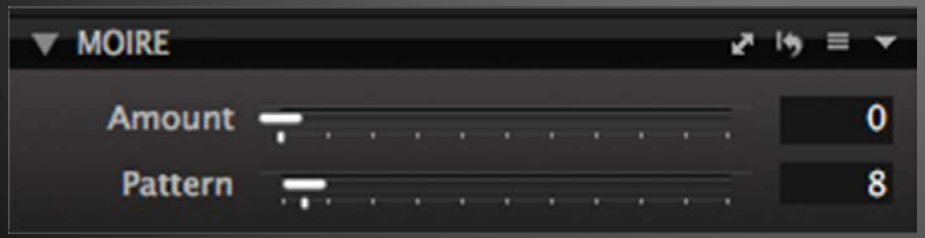
Set Base Characteristics

Assign Capture Adjustments

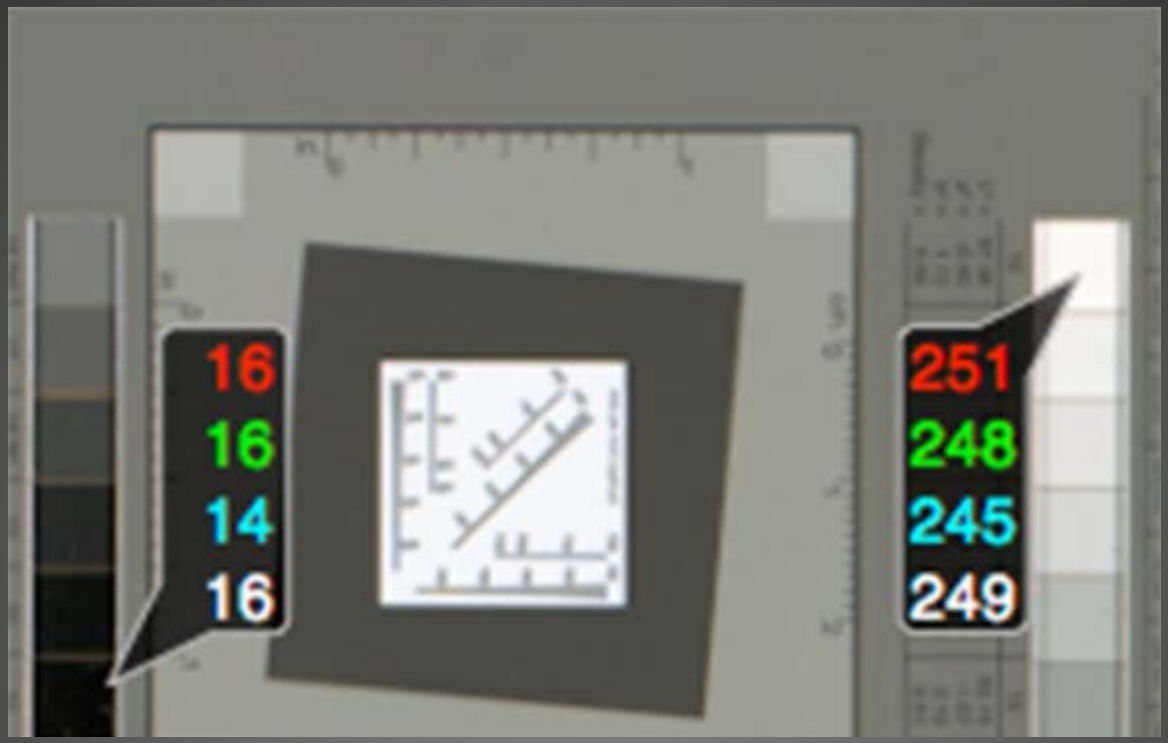


Set White Balance

Smithsonian
National Museum of Natural History



Bonus: Abridged Camera Workflow
Settings in Details Tab



Bonus: Abridged Camera Workflow

Place a series of RGB read outs by clicking various RGB patches on the Device Level and/or Object Level Targets

Smithsonian
National Museum of Natural History

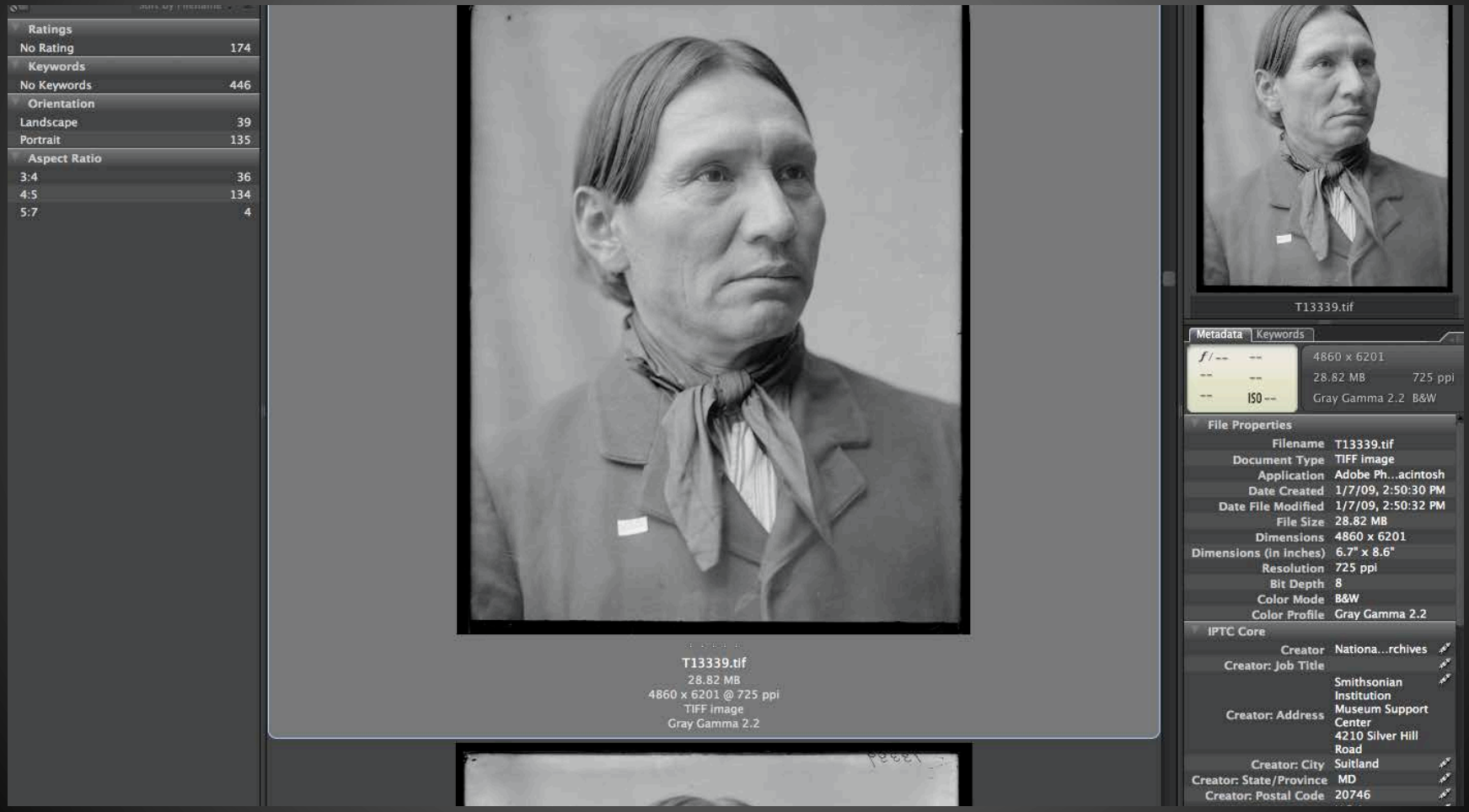
Image Creation – Conclusions

- Considerations:
  - Time
  - Staff
  - Hardware and Software
  - Imaging Environment
  - Bit depth
  - Tone
  - Color Mode
  - Resolution
  - File Formats

Don't forget to use a target!

Smithsonian
National Museum of Natural History

More is More: The Role and Responsibility of Metadata



What is metadata?

- Literally "data about data"
- Descriptive information about collections
- value-added information used to arrange, describe, track, and enhance access to collections

- Metadata is more common in daily online life than you might think:
  - Photo tagging (Facebook, Instagram)
  - #Hashtags (Twitter, Instagram, Facebook)
  - Blog tags

What are the different types of metadata?

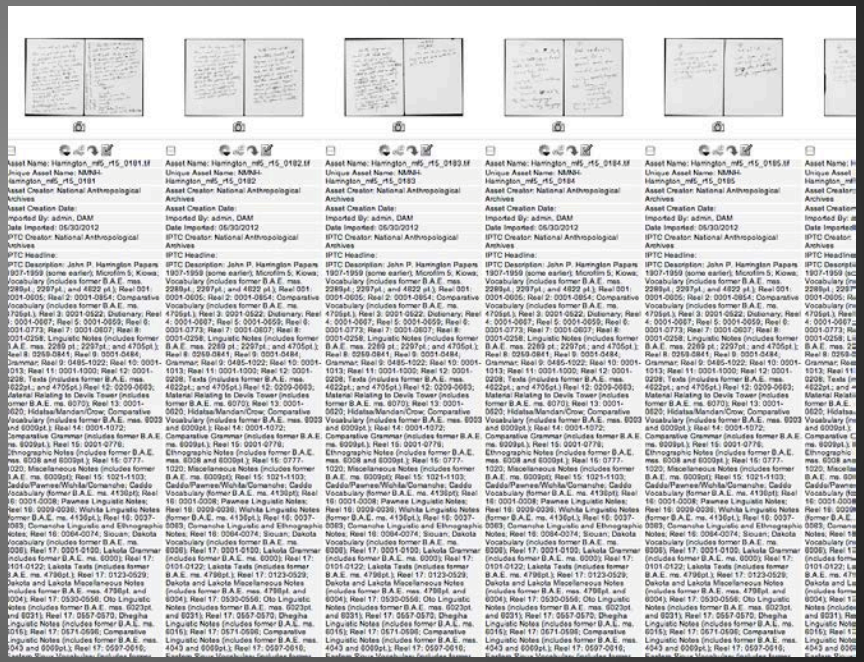
(Source: Source: Baca, Murtha, editor. Gill, Tony, Gilliand, Anne J., Whalen, Maureen, and Woodley, Mary S. (2008). "Setting the Stage," Introduction to Metadata Online Edition, Version 3.0, http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html

Type	Definition	Example
Administrative	Metadata used in managing and administering collections and information resources	- Acquisition information - Rights and Reproduction tracking - Documentation of legal access requirements - Location information - Selection criteria for digitization
Descriptive	Metadata used to identify and describe collections and related information resources	- Cataloging Records - Finding Aids - Differentiations between versions - Specialized indexes - Curatorial information - Hyperlinked relationships between resources - Annotations by creators and users
Preservation	Metadata related to the preservation management of collections and information resources	- Documentation of physical condition of resources - Documentation of actions taken to preserve physical and digital versions of resources - Documentation of any changes occurring during digitization or preservation
Technical	Metadata related to how a system functions or metadata behaves	- Hardware and software documentation - Technical digitization information - Tracking of system response times - Authentication and security data
Use	Metadata related to the level and type of use of collections and information resources	- Circulation records - Physical and digital exhibition records - Use and user tracking - Content reuse and multiversioning information - Search logs - Rights metadata

Smithsonian
National Museum of Natural History

Where is metadata?

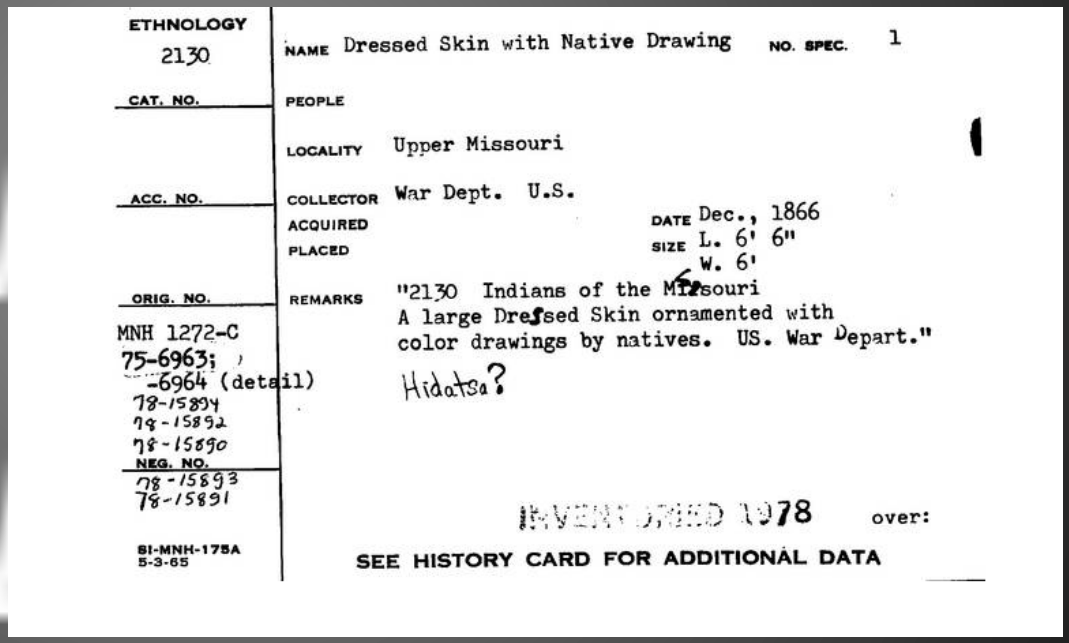
- Different levels of metadata exist in different places
  - Embedded into files
  - Filing cabinets!
    - Accession Records
    - Catalog Cards
  - Collections Information Systems and Databases
    - Online, public-facing platform
    - Internal collections database system



Smithsonian
National Museum of Natural History

When is metadata?

- Collections staff has been working with metadata since the beginning – term not commonly utilized until collections started being digitized
- Metadata should be collected and applied to files as soon as possible – lag time between image creation and metadata enhancement could result in inconsistencies and missing information
- Build metadata into digitization workflow



ETHNOLOGY
2130

CAT. NO.

ACC. NO.

ORIG. NO.
MNH 1272-C
75-6963;
-6964 (detail)
78-15894
78-15892
78-15890

NEG. NO.
78-15893
78-15891

SI-MNH-175A
5-3-65

NAME Dressed Skin with Native Drawing NO. SPEC. 1

PEOPLE

LOCALITY Upper Missouri

COLLECTOR War Dept. U.S.

ACQUIRED

PLACED

DATE Dec., 1866

SIZE L. 6' 6"
W. 6'

REMARKS "2130 Indians of the Missouri
A large Dressed Skin ornamented with
color drawings by natives. US. War Depart."

Hidatsa?

INVENTORIED 1978 over:

SEE HISTORY CARD FOR ADDITIONAL DATA

Who is metadata?

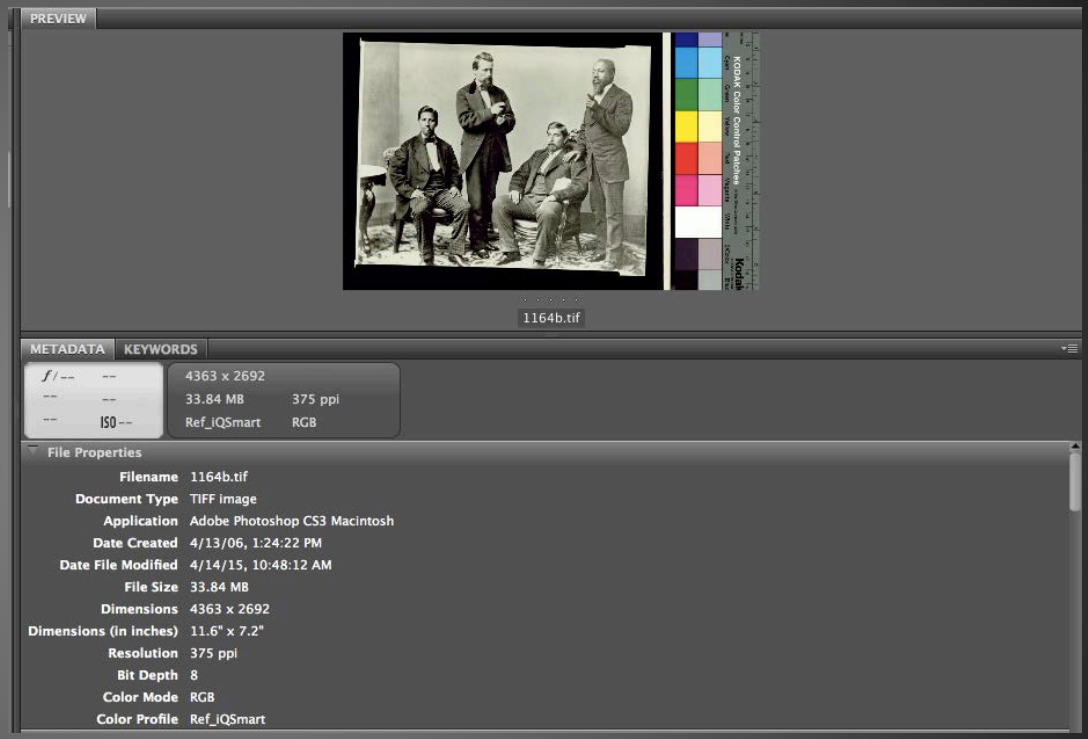
- Standards bodies
  - JEITA
  - IPTC
  - PLUS
- Photographer trade associations
  - SAA
  - ASMP
  - APA
  - EP
  - NPPA
- Cultural Heritage and Government Institutions
  - AAT (Getty)
  - LOC
  - FADGI

- When working with metadata, some "who's" to think about:
  - Who is supplying the metadata?
    - Curators
    - Donors
    - Collections staff (archivists, librarians)
    - Photographers/Imaging Technicians
  - Who is applying the metadata at its various levels?
    - Catalog records
    - Database entries
    - Embedded files
  - Who is using the metadata?
    - Staff
    - Researchers
    - Exhibitions
    - Publications

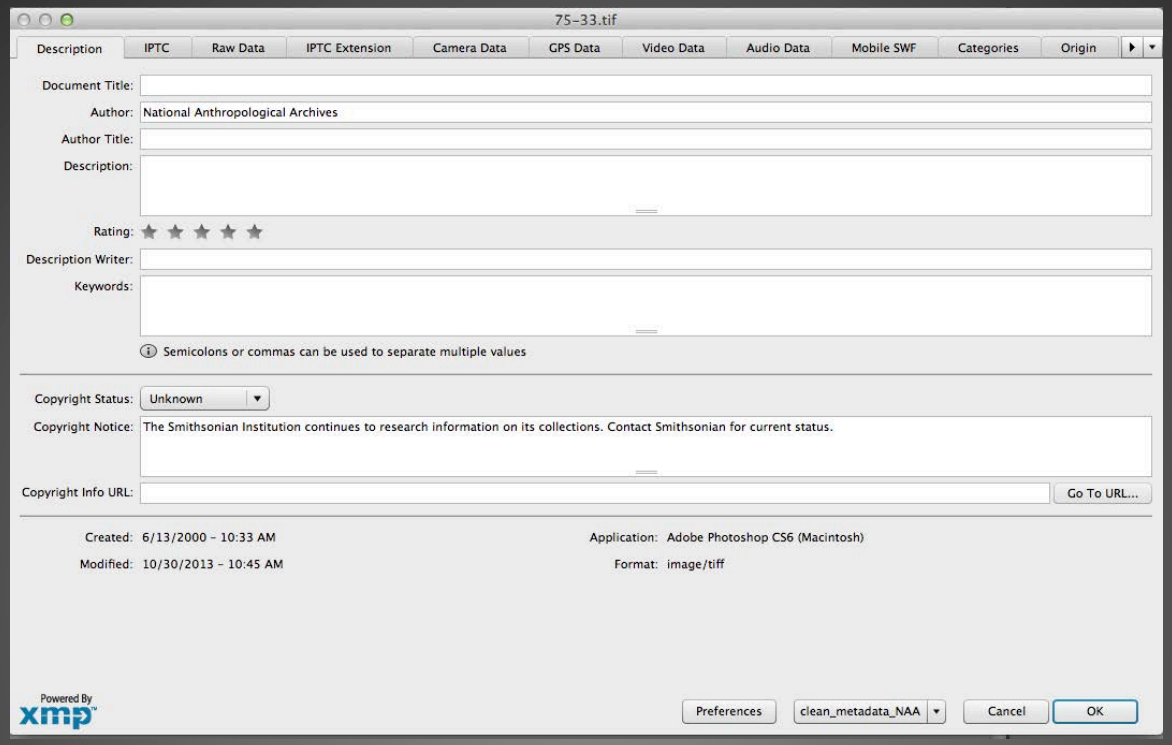
Smithsonian
National Museum of Natural History

Why metadata?

- Accessibility
- Discoverability
- Collections Management
- Usability
- Rights and Restrictions
- Standard Policies and Procedures



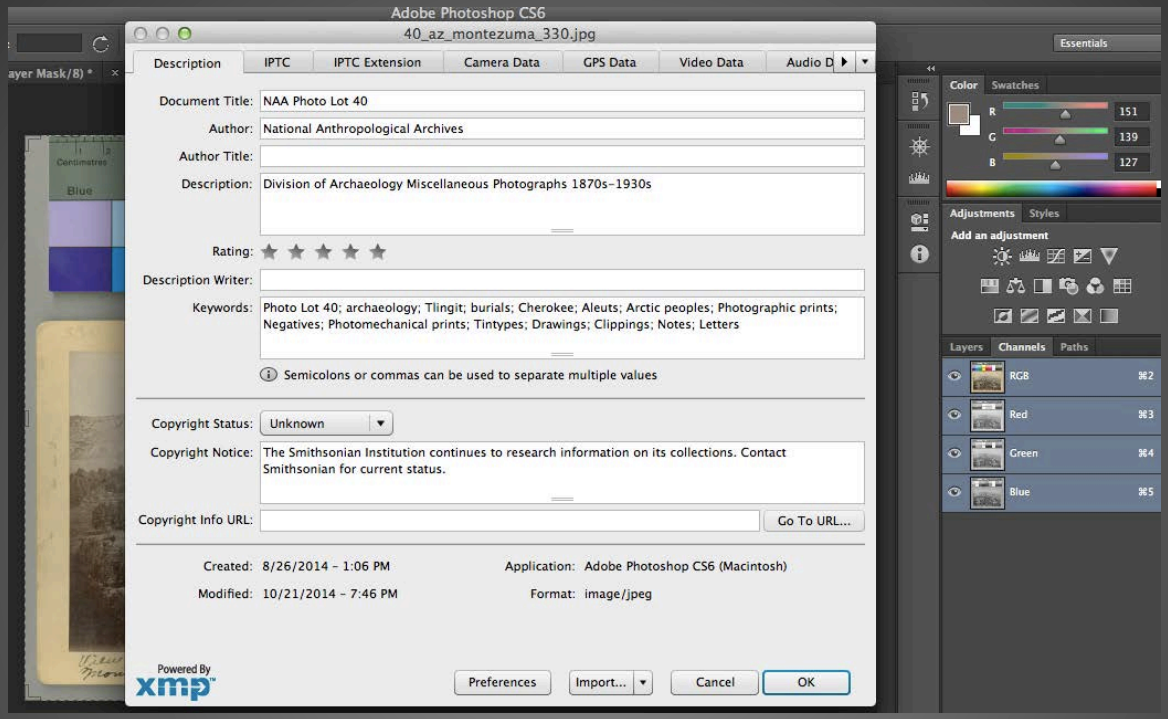
Smithsonian
National Museum of Natural History



How metadata?

- IPTC: International Press Telecommunications Council
- Dublin Core
- Descriptive and Technical
- Embedded

Smithsonian
National Museum of Natural History



How metadata?

Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images
by EMDaWG (Embedded Metadata Working Group – Smithsonian Institution)
April 2010

Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators , April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

Smithsonian
National Museum of Natural History

How metadata? Requirements

Required: Core Set of Embedded Metadata

Element Name	Definition	Sample Data Value	Notes	Character Limit*
Document Title	File number, Accession Number, Catalog Number, Negative Number, Unique Identifier root level etc.	LB016021-a 08596201 gn_03644 landes_photo_arizona_16 digital filename 23456.000 **catalog number* 123457.000;1234568.000 **multiple objects in one image* P00001 **catalog number* N00001 **catalog number* T00001 **catalog number* 20100121_01a_csf_ps_001.tif **file name. represents coverage of event at NMAI* SFF2009_Strauss_6-24_0004.dng		*IPTC fields have character limits depending on the application utilizing the standard and not all data may be displayed. (Add char limit info into the tables) With some applications data may be truncated at a given character limit.
Copyright Notice	Copyright Notice	This image was obtained from the Smithsonian Institution. Unless otherwise noted, this image or its contents may be protected by international copyright laws. **This statement is a default statement and you may use a more specific statement; check with OGC for a specific statement.*	Approved and suggested by OGC.	
Source	Name and Abbreviation of SI owning unit, Smithsonian Institution	NMAI-Natl. Museum of the American Indian, Smithsonian Institution NAA- Natl. Anthropological Archives, Smithsonian Institution		
Creator (*Note: Unit makes decision-document within unit how they reached this decision). In the DAMS the IPTC Creator field is mapped to the Asset Creator field, which is the creator of the digital object.	Creator of digital object *or* Creator of original object	Edward F. Caldwell and Co. Department of Anthropology National Anthropological Archives Photographer Name **If author is not known then default to department name (refrain from using acronyms)* Cynthia Frankenburg **if Creator=name then Creators Title field is populated. Job Title=Photographer* William Greene **if Creator=name then Creators Title field is populated. Job Title=Scanner* Woody Guthrie **Creator original object*	*IPTC Creator Job Title field can be used to define the role of the creator.	

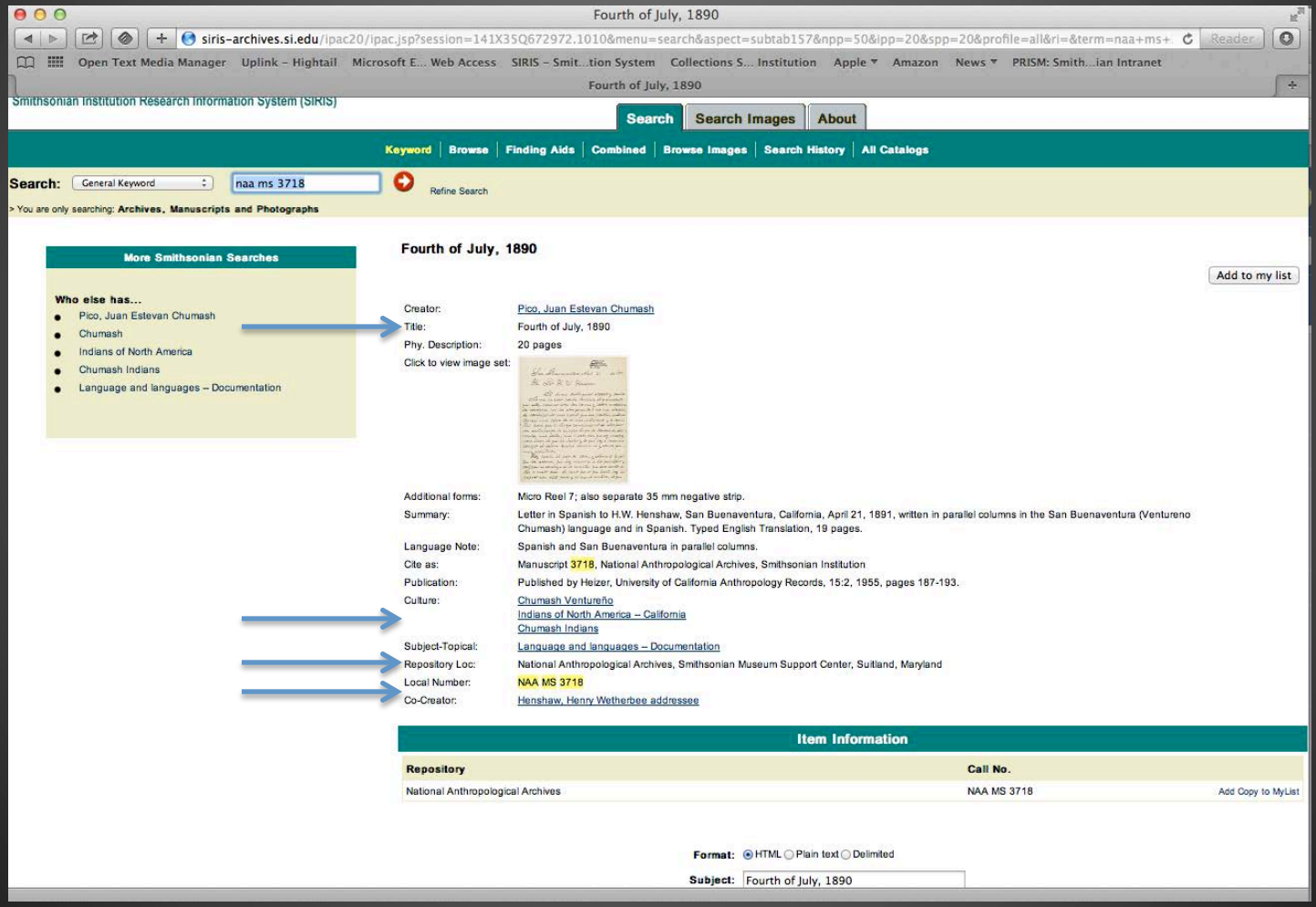
Smithsonian
National Museum of Natural History

How Metadata? Suggested

Element Name	Definition	Sample Value	Notes	Character Limit
Date *In the DAMS IPTC Date is mapped to Asset Creation Date, which is the creation date of the digital object.*	Date of Object *or* Date of Creation of Digital	07/01/1967	Description field can be used if date range or other date structure is being used.	
Description	Free narrative text	123457.000 (right); 123458.000 (left) **see above under Title example. This field is used to locate individual objects within an image that contains multiple objects.* Cultural Resources Center 2007 Powwow Open House, CRC Open House, CRC Exterior, Chief Joseph, Nez Perce		
Keywords	Free text field but should be used to store a list of standard term(s) separated by a common delimiter such as semicolon.	Lighting Archive; Electrification; Lighting; Lighting Fixtures; Architectural History; History of Architecture Object; Publication; Our Lives ** taken from defined look-up list in database: **Object**=image of object in the collection; **Publication**= quality suitable for publication; **Our Lives**=imaged for **Our Lives** exhibition.* Alice Fletcher; Francis La Flesche; 4558 Viento de Agua; plena; bomba rugby; Wales; sports	This list can come from any existing controlled vocabularies like your unit CIS' Iconography lookup list, public resources such as Library of Congress Subject Headings, taxonomic checklists, etc. The goal is to be *consistent* as this is a field whose data is often used for searching. For instance, if you use singular form, stick with singular form, don't alternate between singular and plural. Don't alternate between variations, like US, USA, or United States. *(See below for links to controlled vocabularies for consideration.)	
Credit/Provider	What you would like to accompany the image in a publication. Ex: Image Number, SI owning Unit, Smithsonian Institution	Image Number, National Anthropological Archives, Smithsonian Institution Papers of Ruth Landes, National Anthropological Archives, Smithsonian Insitution Ken Rahaim, Smithsonian Institution ** Ernest L. Spybuck (Absentee Shawnee, 1883–1949), *Procession before War Dance*, ca. 1910. Watercolor on paperboard, 42.2 x 63.9 cm. Oklahoma. Photo by David Heald. 2/5735		
Job Identifier	Instructions or unit id for a job	MSC07-04608		
Headline	(Formally called Caption) A descriptive title or a caption.	Dr. J.E. Tallmage		

Smithsonian
National Museum of Natural History

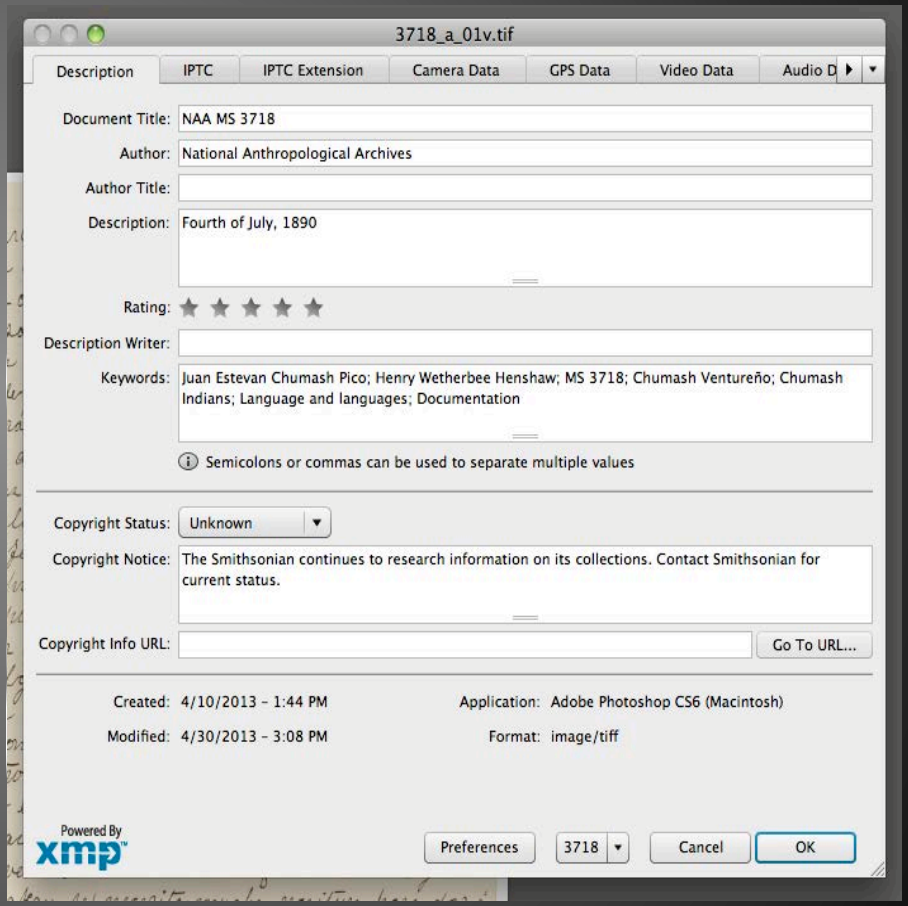
Locating metadata content



Smithsonian
National Museum of Natural History

Creating a metadata template

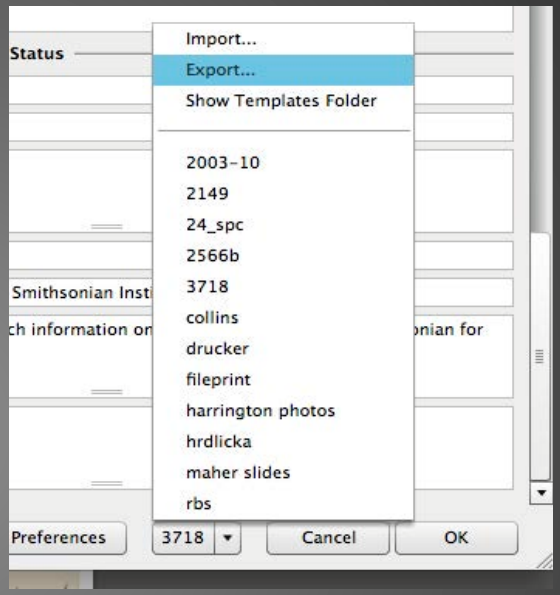
- Open a file from the folder you're working on, and go to File→File Info in Photoshop. Using the EMDAWG guidelines, the following fields are populated:
- Document Title: The number of the manuscript (from the Local Number field in SIRIS).
- Author: National Anthropological Archives (EMDAWG – Creator)
- Description: The title of the manuscript, also from the SIRIS record.
- Keywords: The creator of the manuscript; a co-creator (if one exists); MS and its number; the culture the manuscript pertains to; and the subject-topical phrases. These all come from the SIRIS record
- Copyright Status: UNKNOWN. It is extremely important that the copyright status says UNKNOWN.
- Copyright Notice: "The Smithsonian continues to research information on its collections. Contact Smithsonian for current status." This is the same for all material.
- In the IPTC tab, make sure the Source field is populated. It should say, "NAA-Natl. Anthropological Archives, Smithsonian Institution."

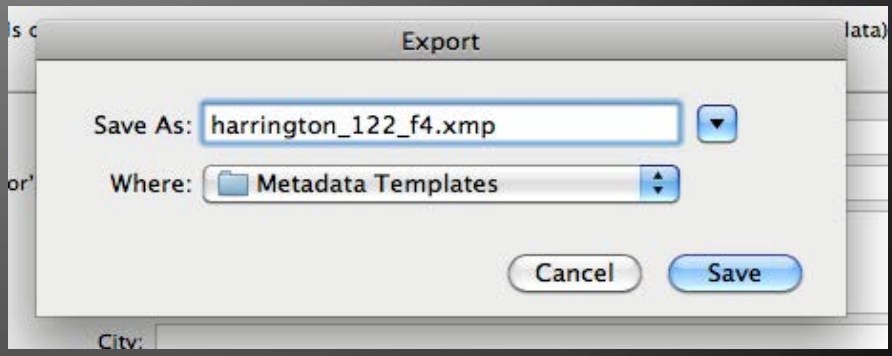


Smithsonian
National Museum of Natural History

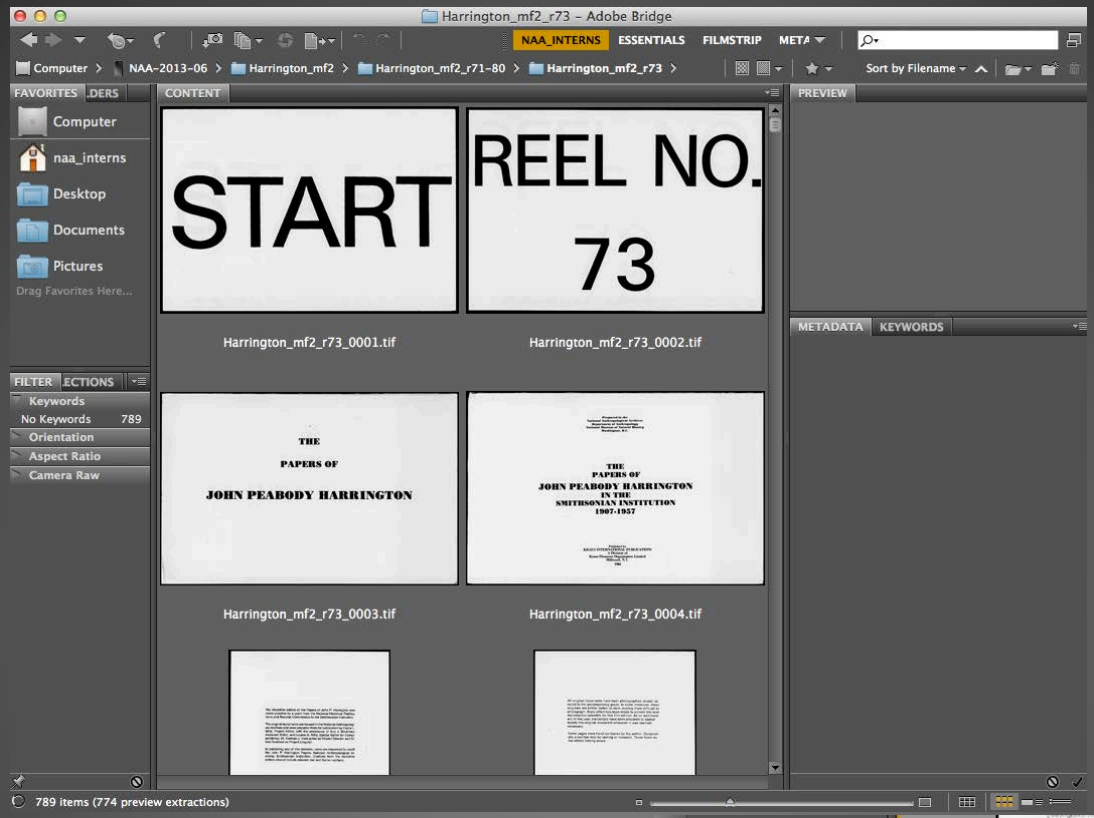
Exporting Metadata

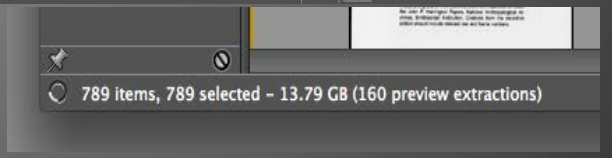
- Once you have created the metadata template, you will want to Save (or Export) it for later use.
- In the metadata window, click Export. A new Save window will pop up, and it will automatically fill in the filename of the file currently open.
- You may leave it as such, or make it shorter and simpler, like just using the number of the manuscript for which the metadata applies.
- Click save. You have created a metadata template that can be used on a collection of images as needed.





Smithsonian
National Museum of Natural History





Embedding Metadata

Metadata can be embedded more quickly embedded using Adobe Bridge

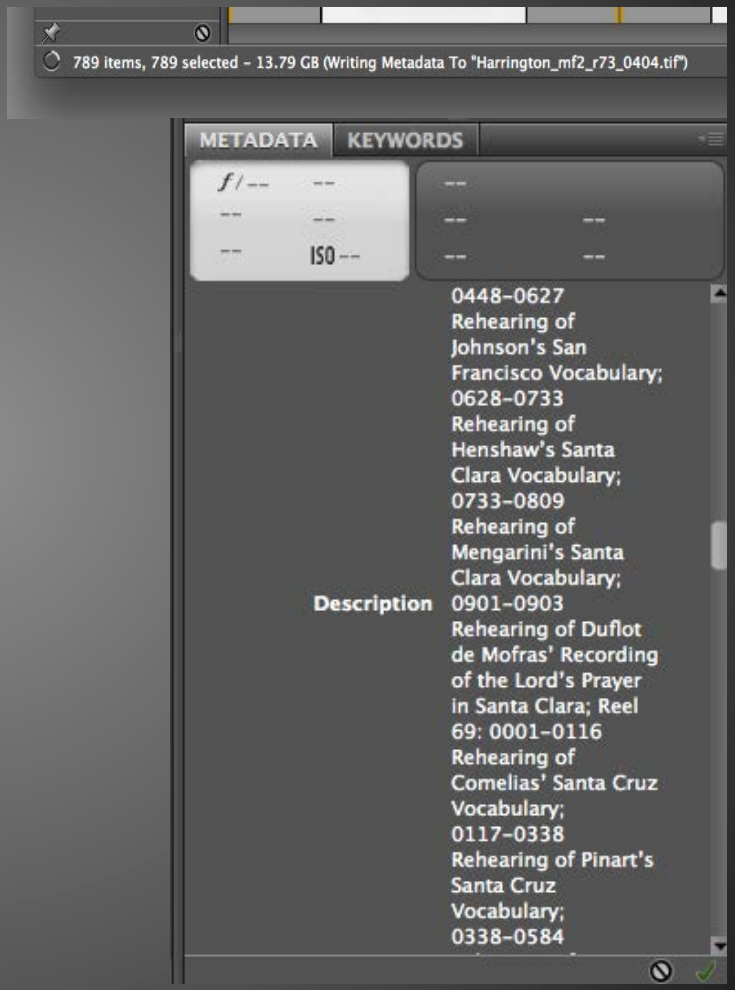
Open the files in Adobe Bridge. The files may take several minutes to load and go through the preview extraction.

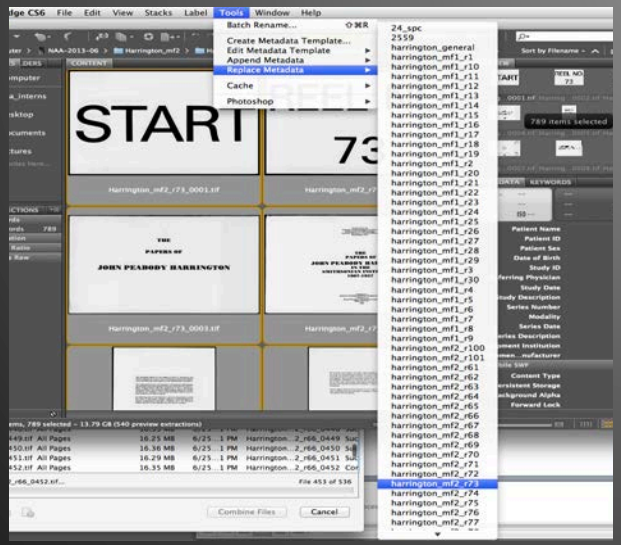
Once all the files have loaded, select all files (command + A) then go to Tools → "Replace Metadata"

Smithsonian
National Museum of Natural History

Embedding Metadata

- Select the corresponding metadata template. Embedding will begin automatically and may take a few moments to complete. Once completed, bridge will automatically save the metadata.
- On the lower right hand side, confirm that the metadata has been embedded.
- Confirm that metadata has been embedded by opening the files in Adobe Photoshop and checking the "File Info."





Metadata - Conclusions

- Metadata is already an integral part of Collections Management
- Embedding and Utilizing Metadata will increase accessibility, usability, discoverability, organization, alleviate rights and restrictions issues

Smithsonian
National Museum of Natural History

Quality Control and Post-Processing

How much is too much? How much is enough?

Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators , April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

What is Quality Control?

- Series of processes done after digitization to ensure you are meeting your predetermined standards and benchmarks, while maintaining accuracy, quality, and consistency.

Smithsonian
National Museum of Natural History

Quality Control Checklist

Kenney, Anne R. and Oya Y. Reiger. "Establishing a Quality Control Program," Chapter 4, pp.61-73. (2000). Moving Theory into Practice: Digital Imaging for Libraries and Archives. Mountain View, CA: Research Libraries Group.

Prerequisites for QC

- Identify products and goals
- Agree on Standards
- Determine a Reference Point
- Understand the Limitations of Current Knowledge, Practice and Technology

Setting up your QC Program

- Identify Scope
- Determine Methods
- Evaluate System Performance
- Codify Inspection Procedures
- Control QC Environment
  - Hardware configuration
  - Image-display software
  - Monitor setup
  - Color quality control
  - Color management
  - Viewing conditions
  - Human characteristics

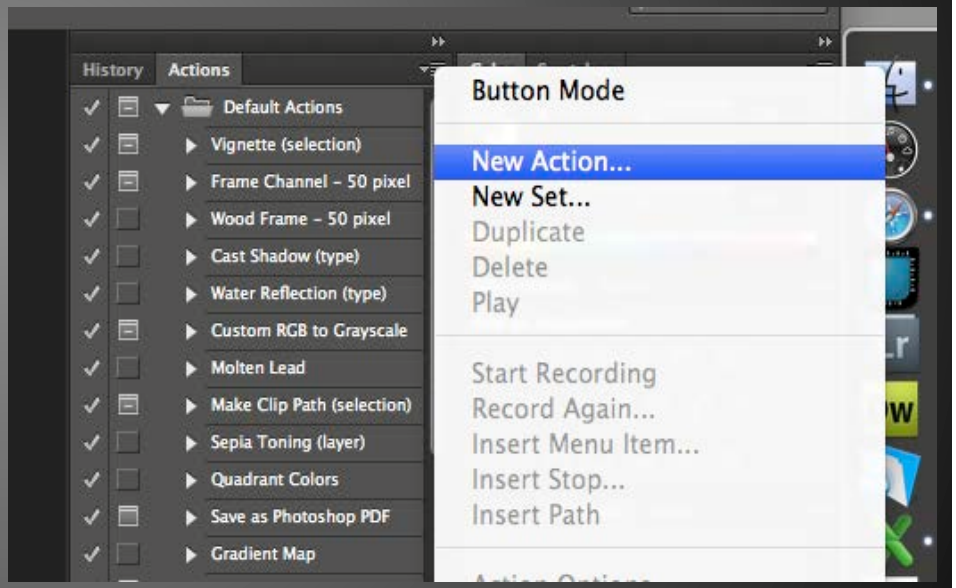
Assessing Image Quality

- Evaluate:
  - Resolution
  - Color and Tone
  - Overall appearance

Photoshop - Actions

- Post-processing needs like conversion (16-bit to 8-bit), rotation, embedding metadata, curve adjustments can be performed in batch actions in Photoshop.

Open Photoshop. Go to the Action tab on the right side and click the dropdown arrow. Click "New Action."

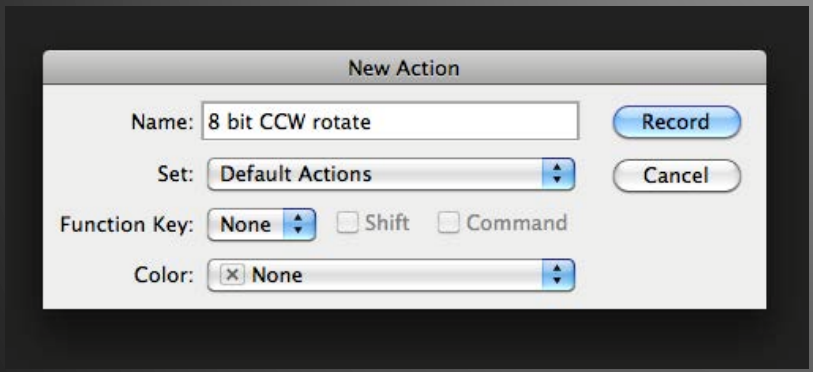


Smithsonian
National Museum of Natural History

Creating an Action

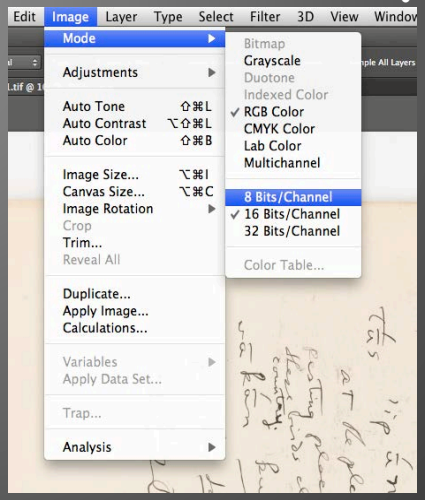
Naming the Action

- A new naming window will pop up. Create a simple, explanatory name for the action, such as "8 bit and CCW rotate."
- Click "record." Photoshop will record all of the steps you complete and save them as a single action that can be applied to an entire folder of images for faster, more efficient processing.

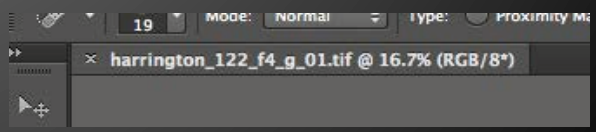


Running the Action

- Go to File → Open and open the first image in the folder you are processing.
- Go to Image→Mode→8 Bits/ Channel



- Once you have converted the image to 8 Bits/ Channel, you may notice that Photoshop will reflect that in the tab where the filename of the image is (RGB/8* indicates this):

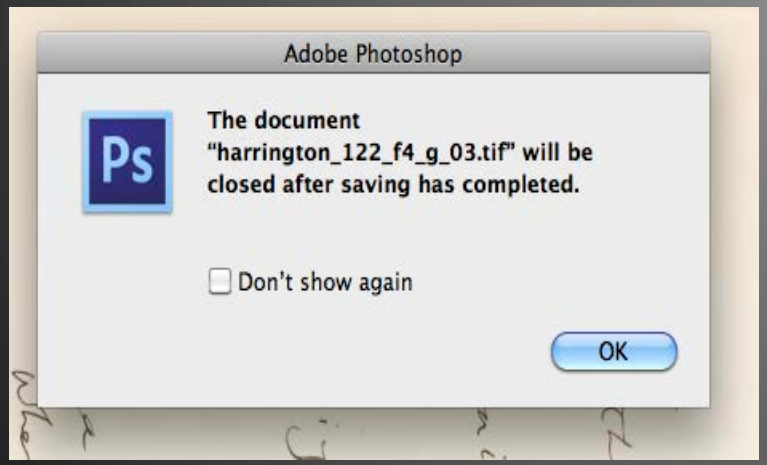


Smithsonian
National Museum of Natural History

Completing an Action

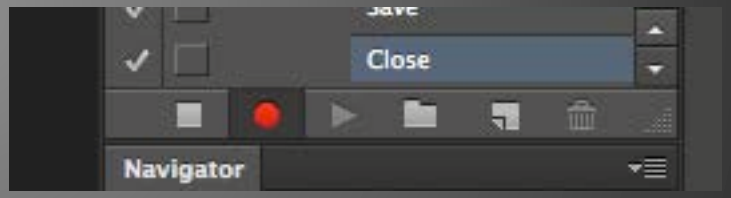
Saving the Changes

- Go to File → Save. Once the dropdown menu has disappeared, go to File → Close. If you get the following popup window, just click Ok.



Stopping the Action

- After you have completed all the steps for the action, you will have to stop recording the action. Click the "stop button," the small square at the bottom of the Action menu.

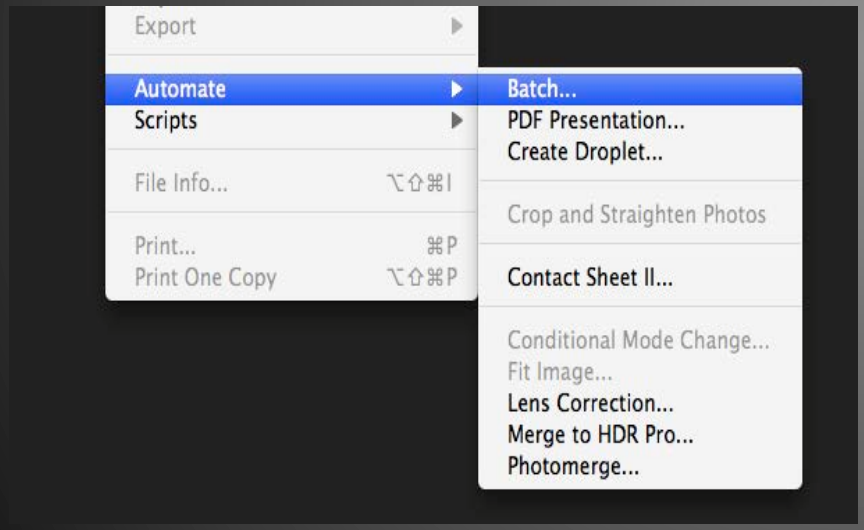


Smithsonian
National Museum of Natural History

Running an Action

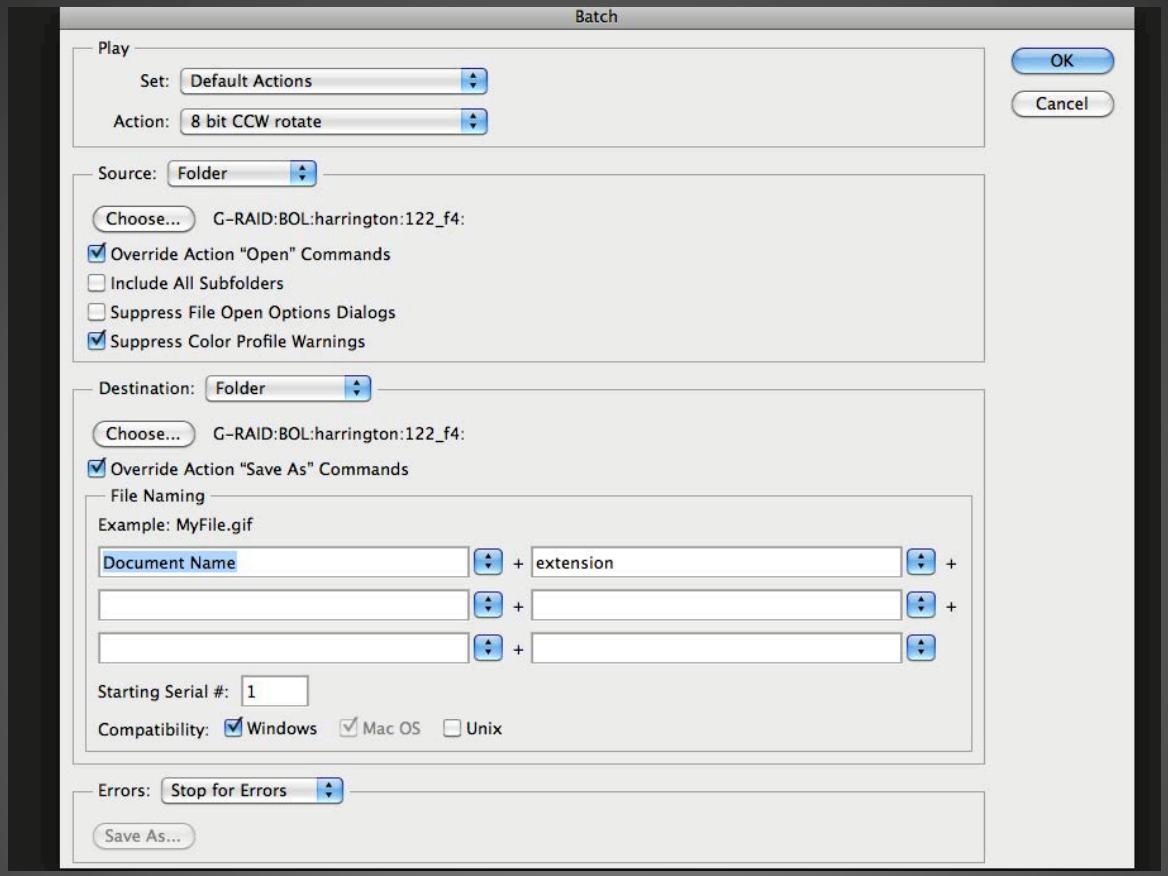
Playing the Action

- Now that you have created the Action, you can "play" it so the steps are applied to the entire folder of images. Go to File → Automate→ Batch



Setting the Action

- A new window called "Batch" will pop up. In the "Play" section select the action you have just created. In the "Source" section click on Choose; locate the folder containing the files you want to run the batch action on; select it. In the Destination section (where the files will end up), do the same as the step before. The window should look like this:

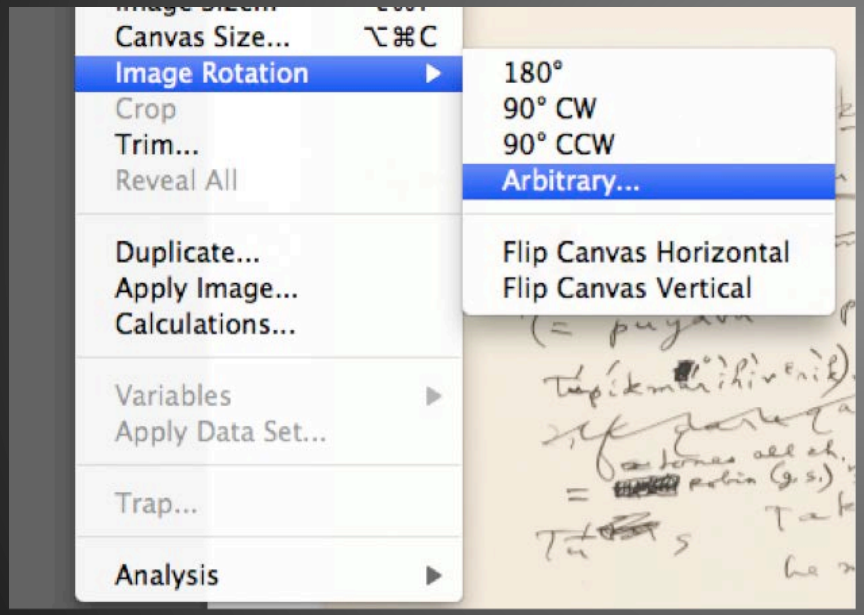


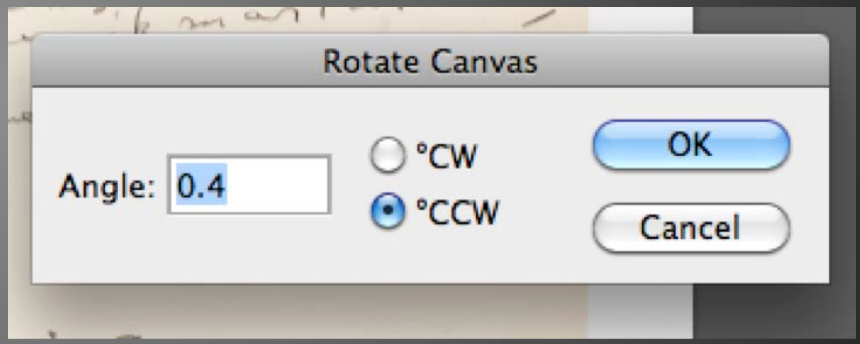
Completing an Action

- Click "Ok." The batching will automatically begin and run through the files contained within the folder, which will take anywhere from a few seconds to a few minutes, depending on the amount and size of the files.

Smithsonian
National Museum of Natural History

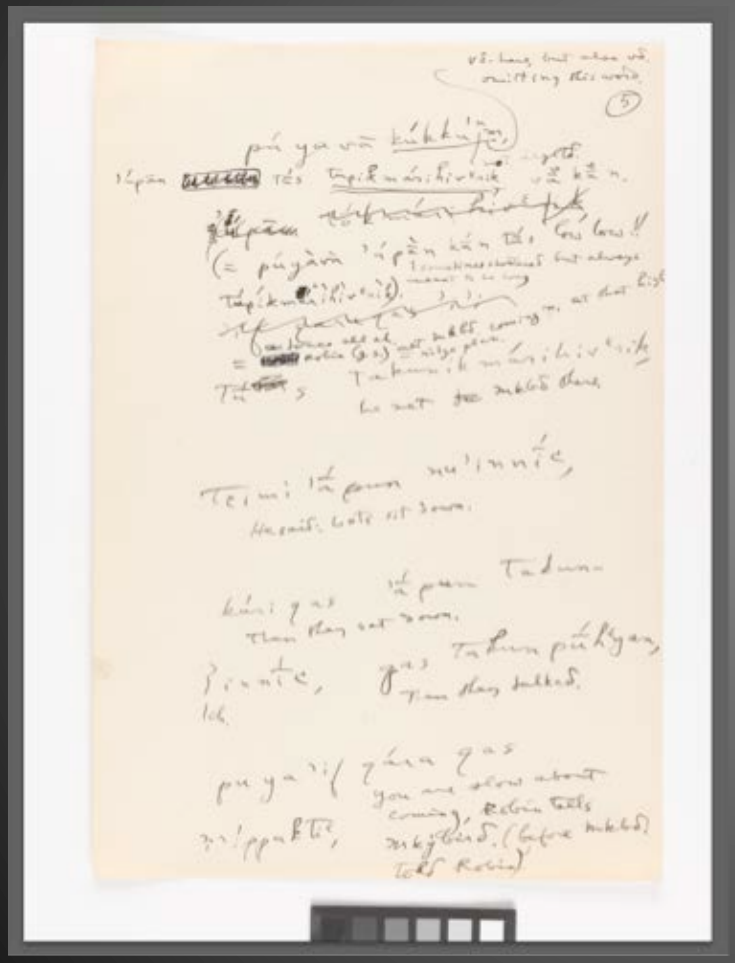
De-skew images as needed

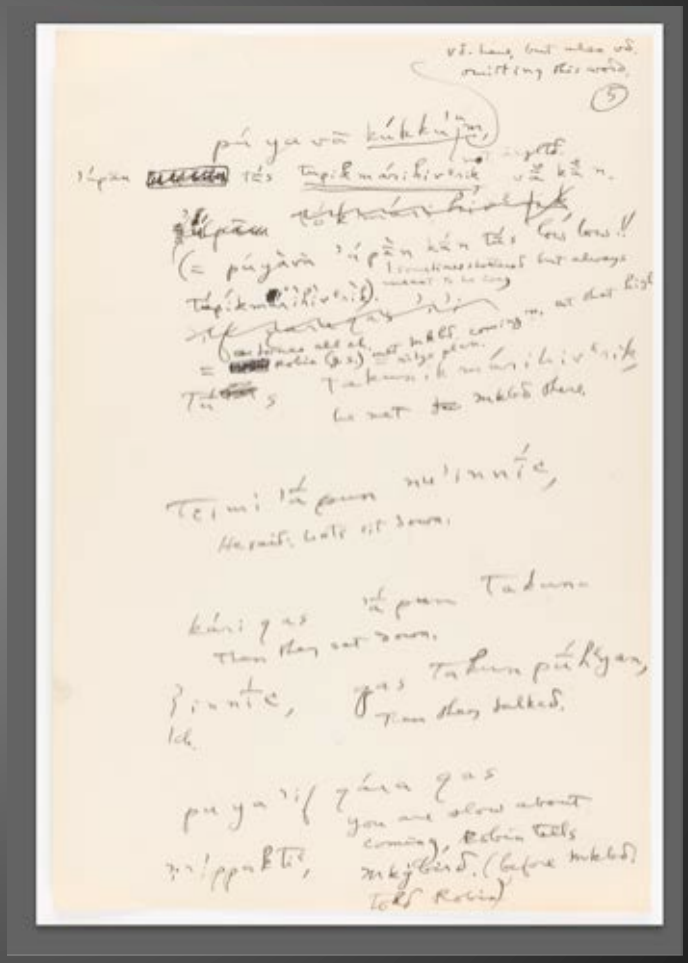




Smithsonian
National Museum of Natural History

Cropping Images

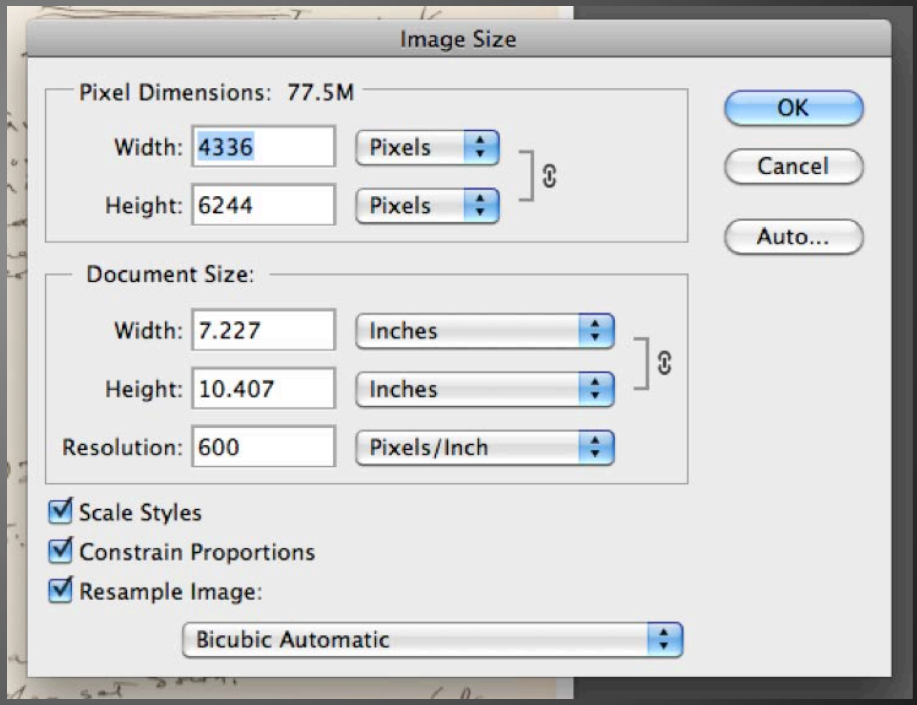




Smithsonian
National Museum of Natural History

Cropping Images

- Once you have cropped the first image, go to Image→Image Size and note the pixel dimensions. You will want to make a note of them so you can set the crop area to these proportions so all the images in a given collection will be the same size.



Smithsonian
National Museum of Natural History

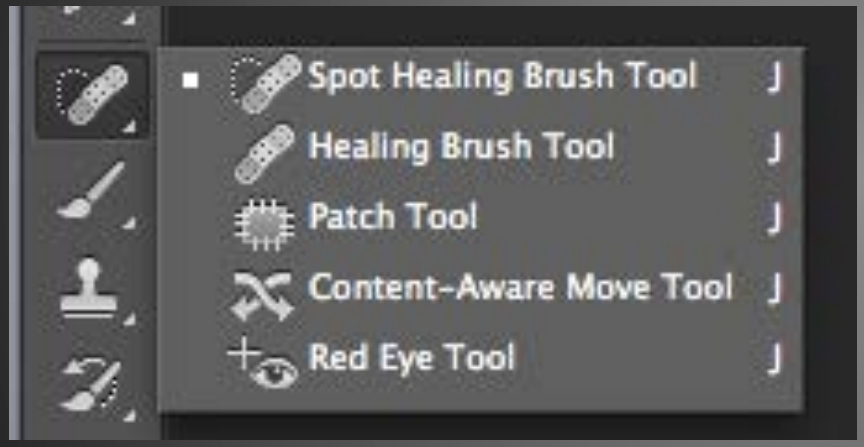
Cropping Images

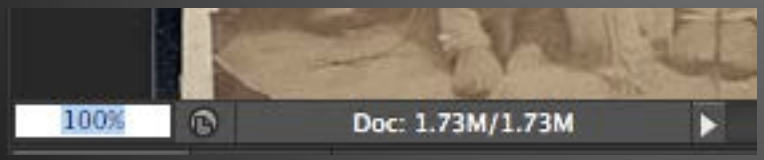
Instead of cropping it "freehand" you will set the crop area based on the pixel information from the previous image.

When you go to crop the image, you will notice that the crop area is constrained to the proportions you've set. Move the crop area around so it fits snugly around the image, while still leaving some background. Once you have cropped it to the correct size, hit enter. The image will be cropped. Click File→Save and File→Close.

Smithsonian
National Museum of Natural History

Inspection of 10% of images at 100% magnification



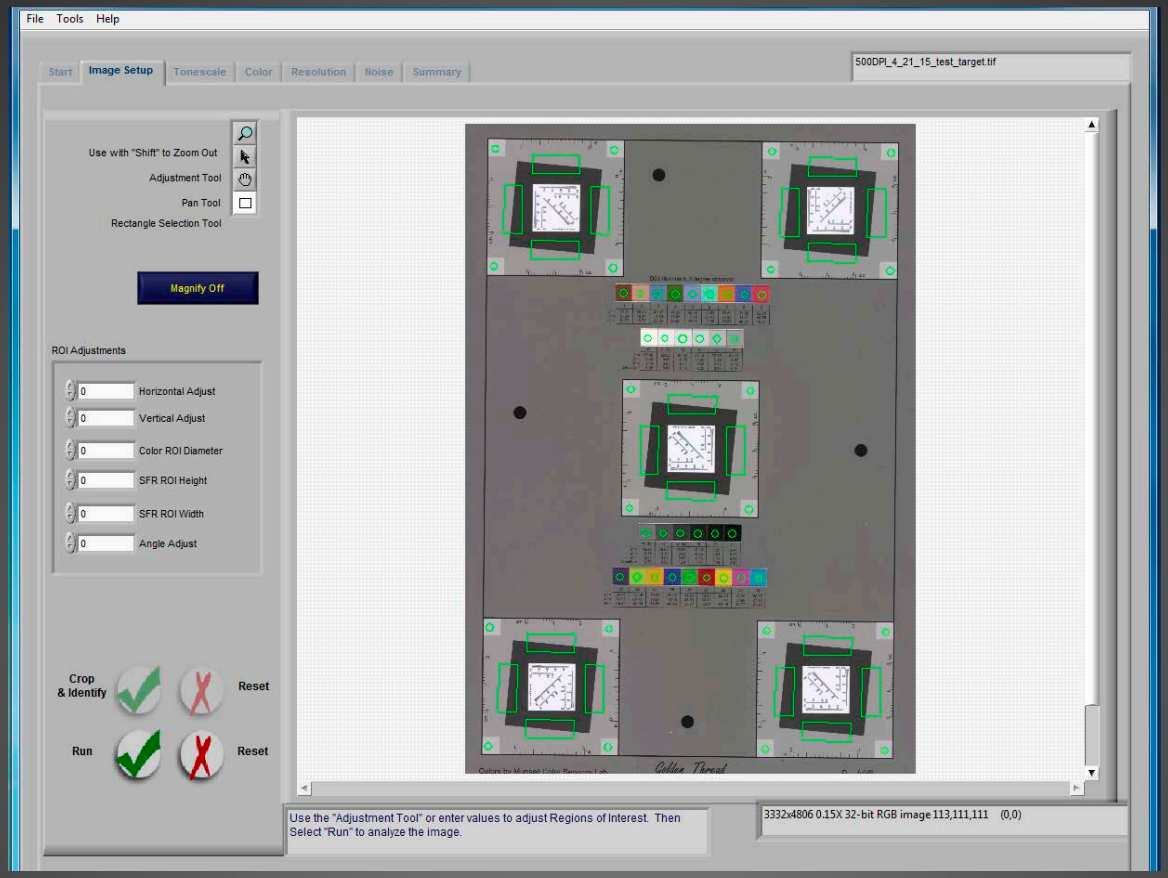


Smithsonian
National Museum of Natural History

Advanced Quality Control

DICE (Digital Image Conformance Evaluation)

Smithsonian
National Museum of Natural History



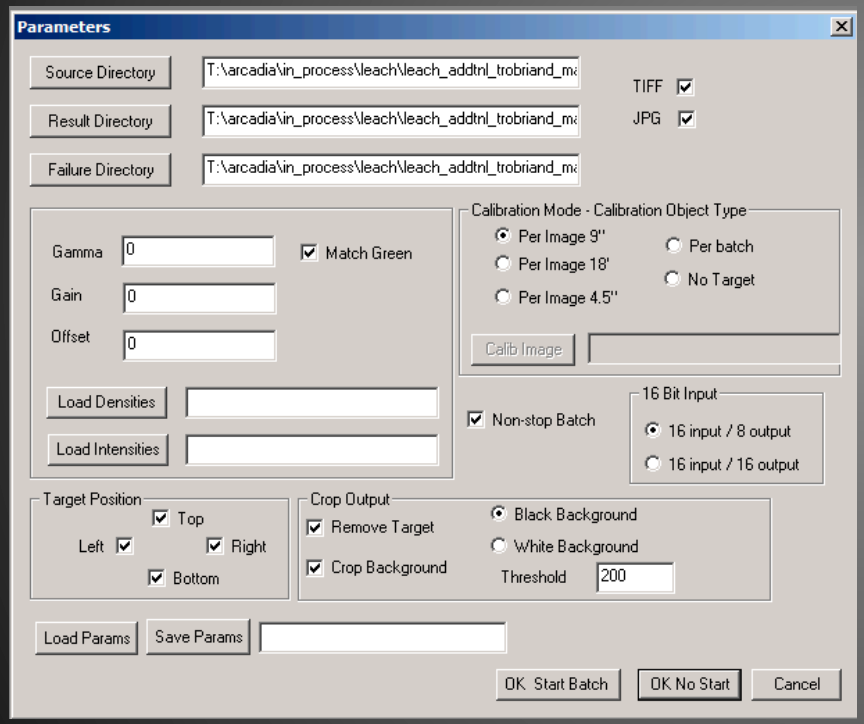
Advanced Quality Control

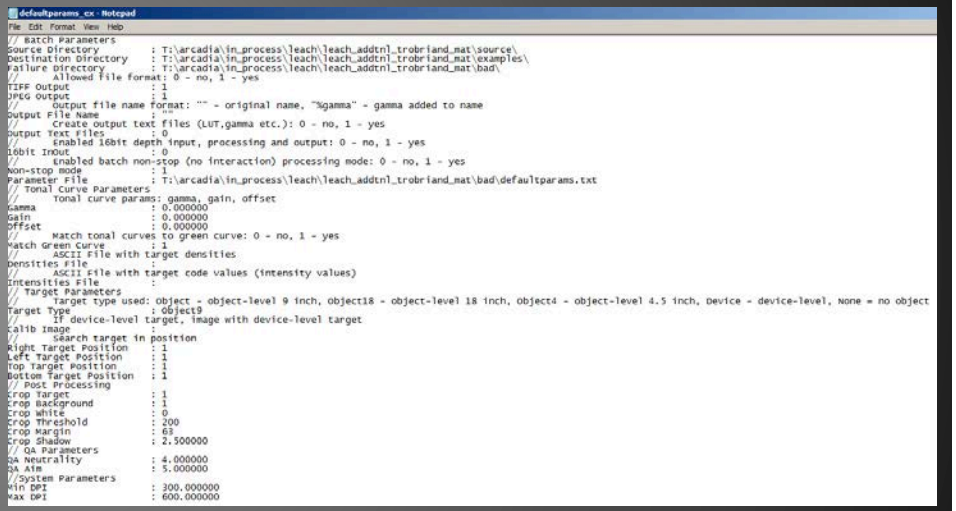
DICE (Digital Image Conformance Evaluation)

Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators , April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

Smithsonian
National Museum of Natural History

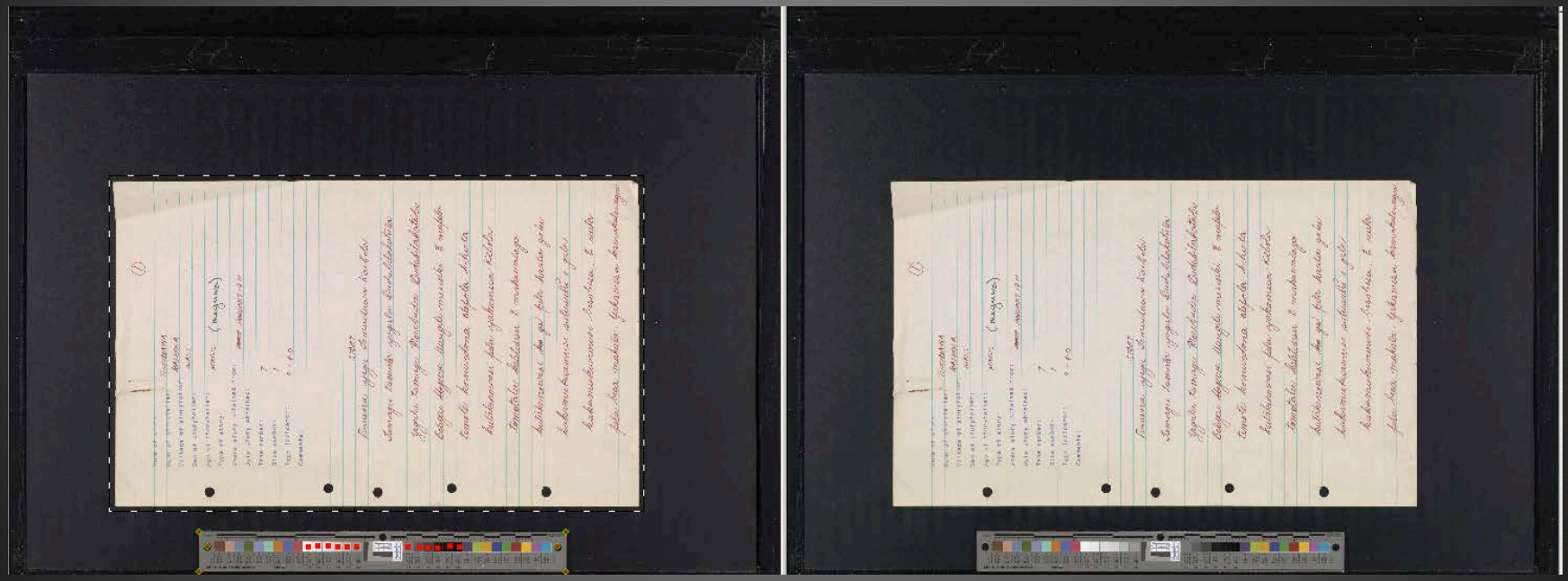
Advanced Quality Control: GoldenTouch Auto-Cropping Software





Smithsonian
National Museum of Natural History

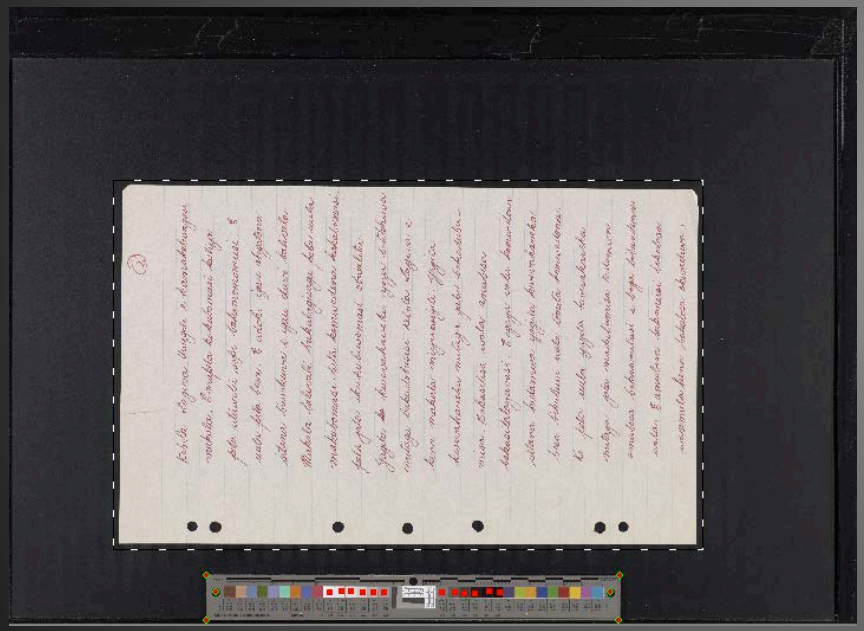
Advanced Quality Control: GoldenTouch Auto-Cropping Software



Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators , April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

Smithsonian
National Museum of Natural History

Advanced Quality Control: GoldenTouch Auto-Cropping Software



Convening Great Lakes Culture Keepers: A Regional Institute for Tribal Librarians, Archivists, and Museum Curators ,
April 26-29, 2015 at the Mille Lacs Indian Museum and Trading Post

Smithsonian
National Museum of Natural History

Quality Control Checklist

Kenney, Anne R. and Oya Y. Reiger. "Establishing a Quality Control Program," Chapter 4, pp.61-73. (2000). Moving Theory into Practice: Digital Imaging for Libraries and Archives. Mountain View, CA: Research Libraries Group.

Prerequisites for QC

- Identify products and goals
- Agree on Standards
- Determine a Reference Point
- Understand the Limitations of Current Knowledge, Practice and Technology

Setting up your QC Program

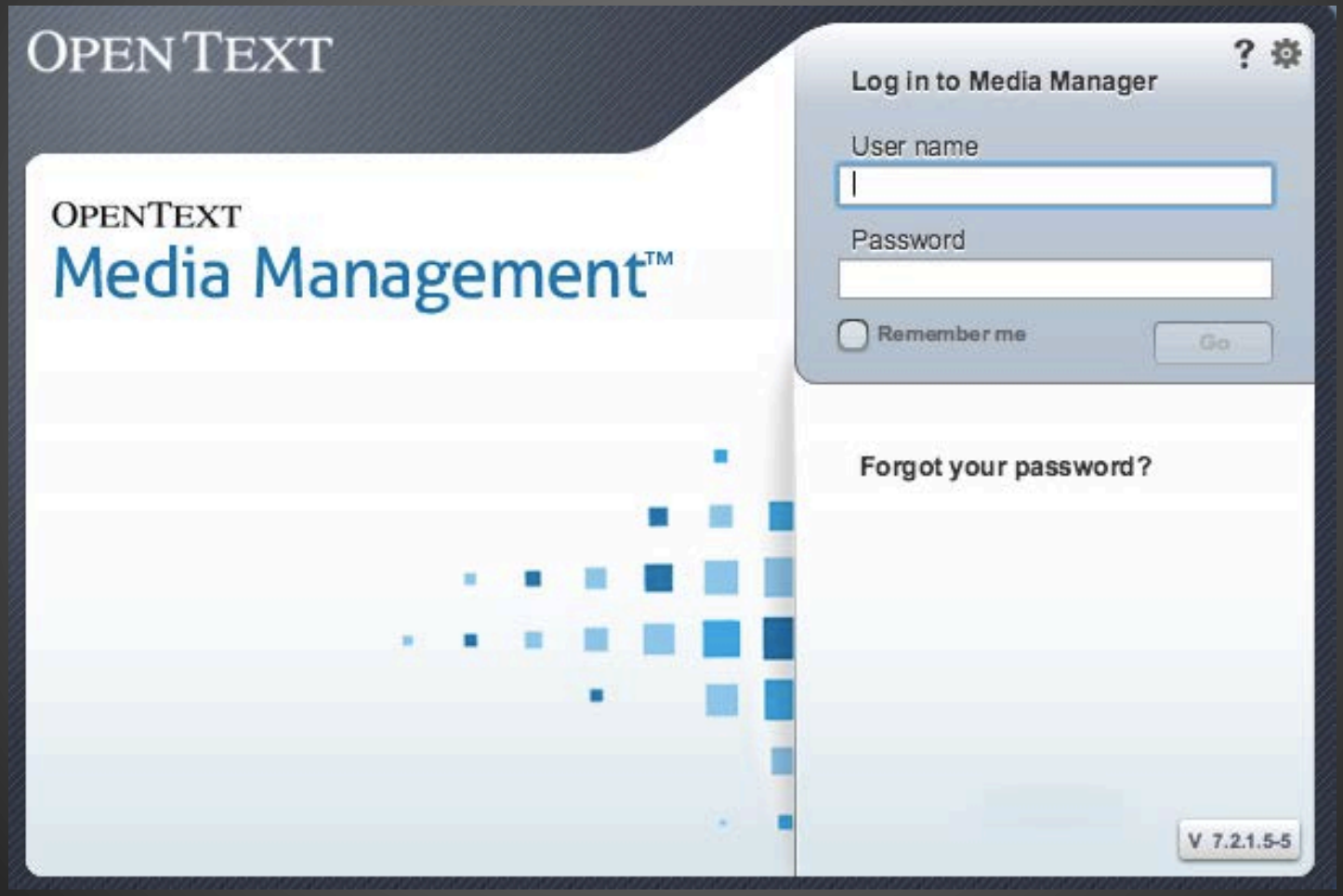
- Identify Scope
- Determine Methods
- Evaluate System Performance
- Codify Inspection Procedures
- Control QC Environment
  - Hardware configuration
  - Image-display software
  - Monitor setup
  - Color quality control
  - Color management
  - Viewing conditions
  - Human characteristics

Assessing Image Quality

- Evaluate:
  - Resolution
  - Color and Tone
  - Overall appearance

Smithsonian
National Museum of Natural History

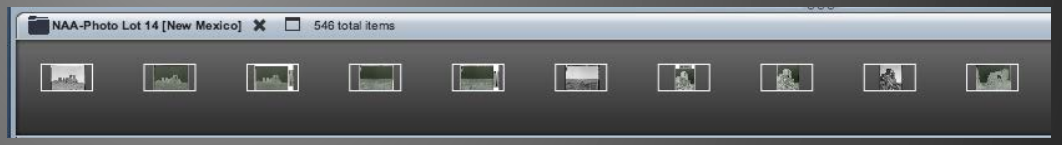
DigitalAssetManagement, Storage!



Strength of Storage

- What are your storage needs?
  - Throughput
- What are your storage capabilities?
  - Staff
  - Budget
- What is the level of IT support?
  - Infrastructure

```
---------- HOT FOLDER IMPORT SUMMARY ----------
Total files to import : 1114
Number of files imported : 1114
Number of files not imported: 0
Number of warnings : 0
-----------------------------------------------------------------------------
```



Smithsonian
National Museum of Natural History

STORAGE WARNING

- DIGITAL FILES NEED TO BE BACKED UP IN MULTIPLE LOCATIONS
- DIGITAL FILES NEED TO BE MIGRATED
- DIGITAL FILES NEED SPACE TO LIVE

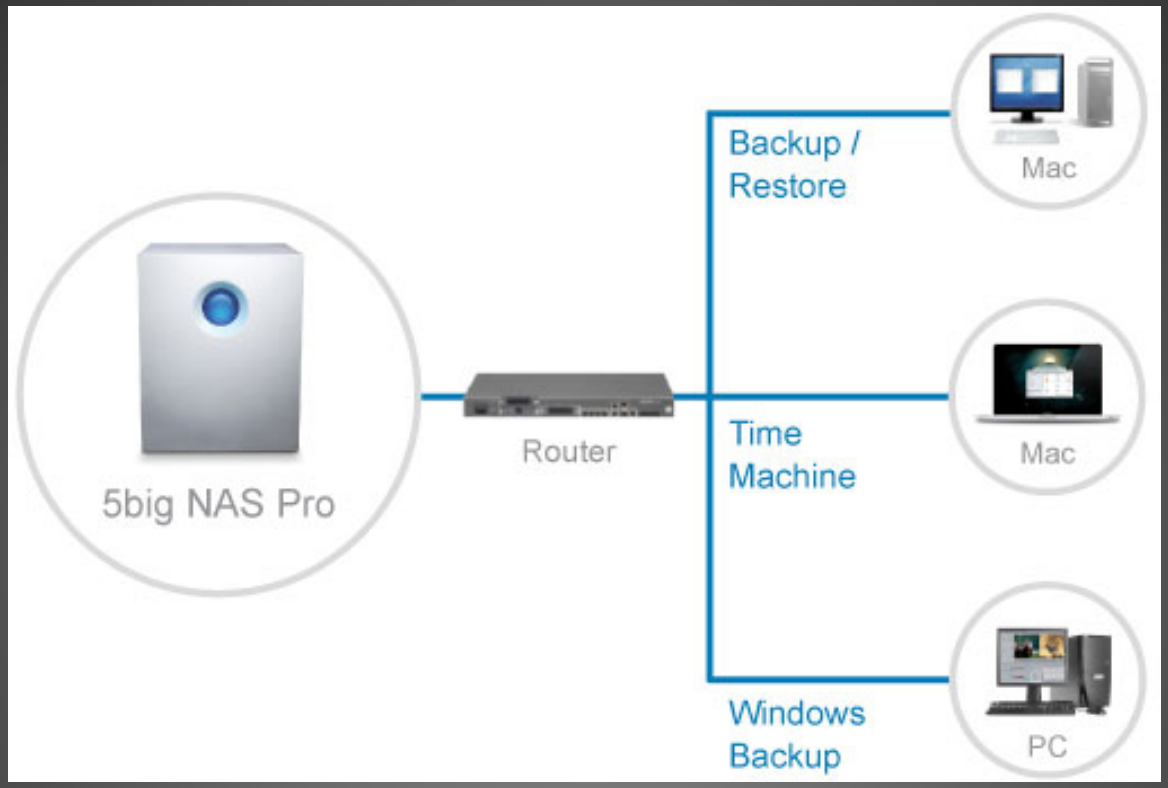
Storage Options

Good, Better 

- Cloud-based Storage
- Network Attached Storage (NAS)
- Locally managed servers
- Local external storage
  - RAID Hard Drives

Best

- Trusted Digital Repositories
- Digital Asset Management Systems
  - Desktop
  - Workgroup
  - Enterprise



Local Storage

Requires constant maintenance and upgrading
Hard drives fail!!
Migration and offsite backup

Smithsonian
National Museum of Natural History

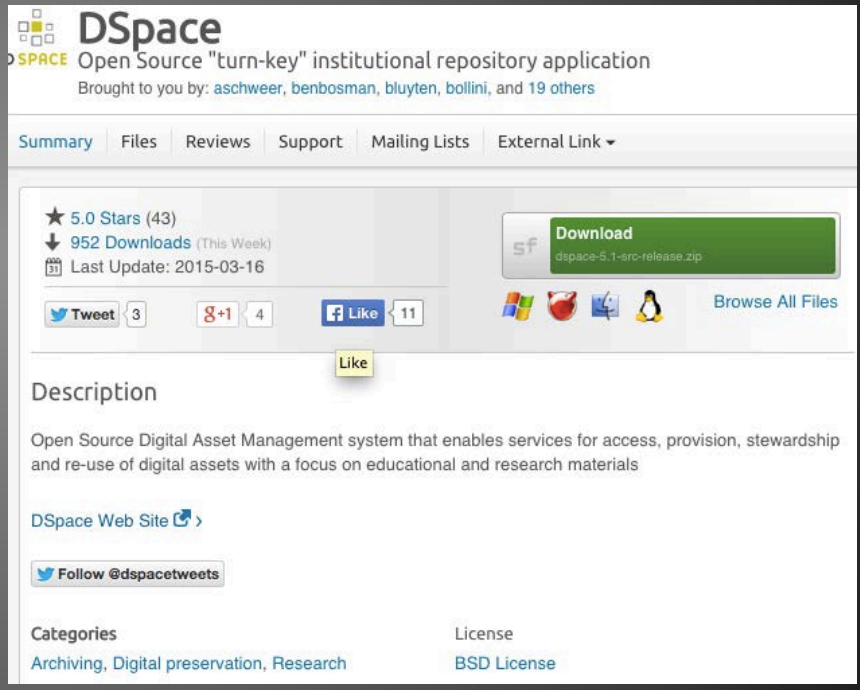
Cloud Storage

- Institutional
  - DuraCloud
  - Chronopolis
  - Digital Preservation Network
- Commercial
  - Amazon
  - Mozy
  - Google Drive
  - iCloud

Smithsonian
National Museum of Natural History

DSpace

- Free, open-source software to build open digital repositories



Smithsonian
National Museum of Natural History

Trusted Digital Repositories

- *"A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future."* (Trusted Digital Repositories: Attributes and Responsibilities An RLG-OCLC Report)
- Digital Preservation Coalition
- HathiTrust (libraries)

DAM useful

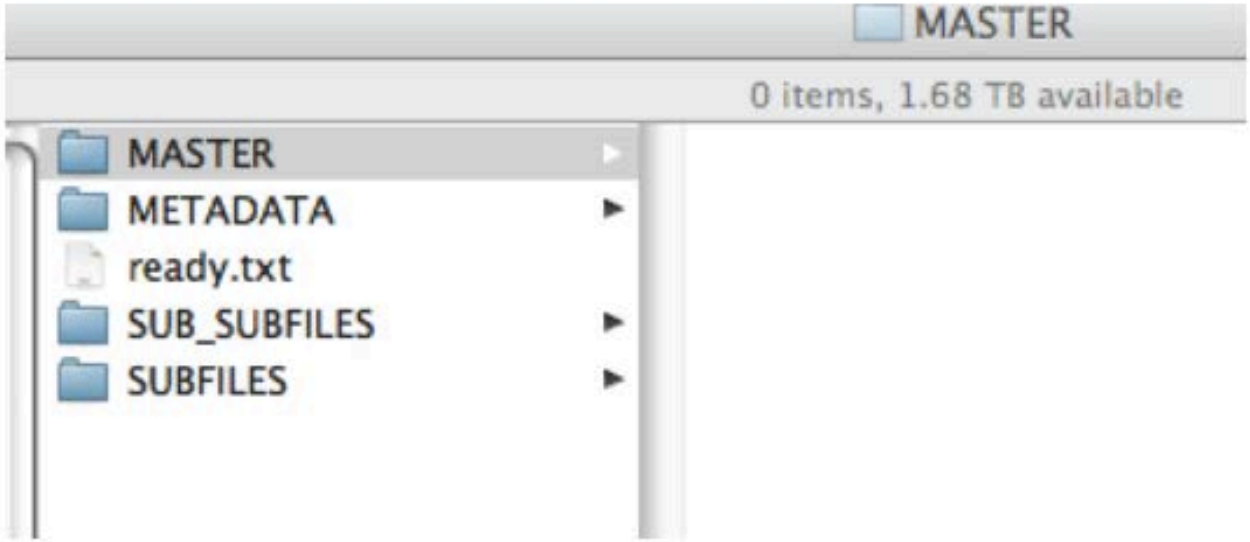
- *"Instituting digital asset management (DAM) in cultural heritage institutions tends to be a major IT initiative. What is often overlooked is that DAM in isolation merely provides a resource intensive organizational tool for digital assets within the institution. The more important aspects of a DAM implementation are the workflow processes and procedures that are integrated into the application, the links the application makes internally and externally to other institutional systems, and ultimately how the implementation changes and enhances the institution's business processes* surrounding the use of digital assets."
  - Howard Goldstein (Center for Digital Imaging, Inc.) and Rob Hendriks (The Rijksmuseum, Amsterdam)

Smithsonian
National Museum of Natural History

DAMS workflow - 1

B. Upload digital assets to the DAMS through the Hot Folder.

1. Sign into the Hot Folder [Go → Connect to Server]. You will see the following folder structure:



2. You will drag files directly into the Hotfolders. Where you copy the files will depend on what you have:

Smithsonian
National Museum of Natural History

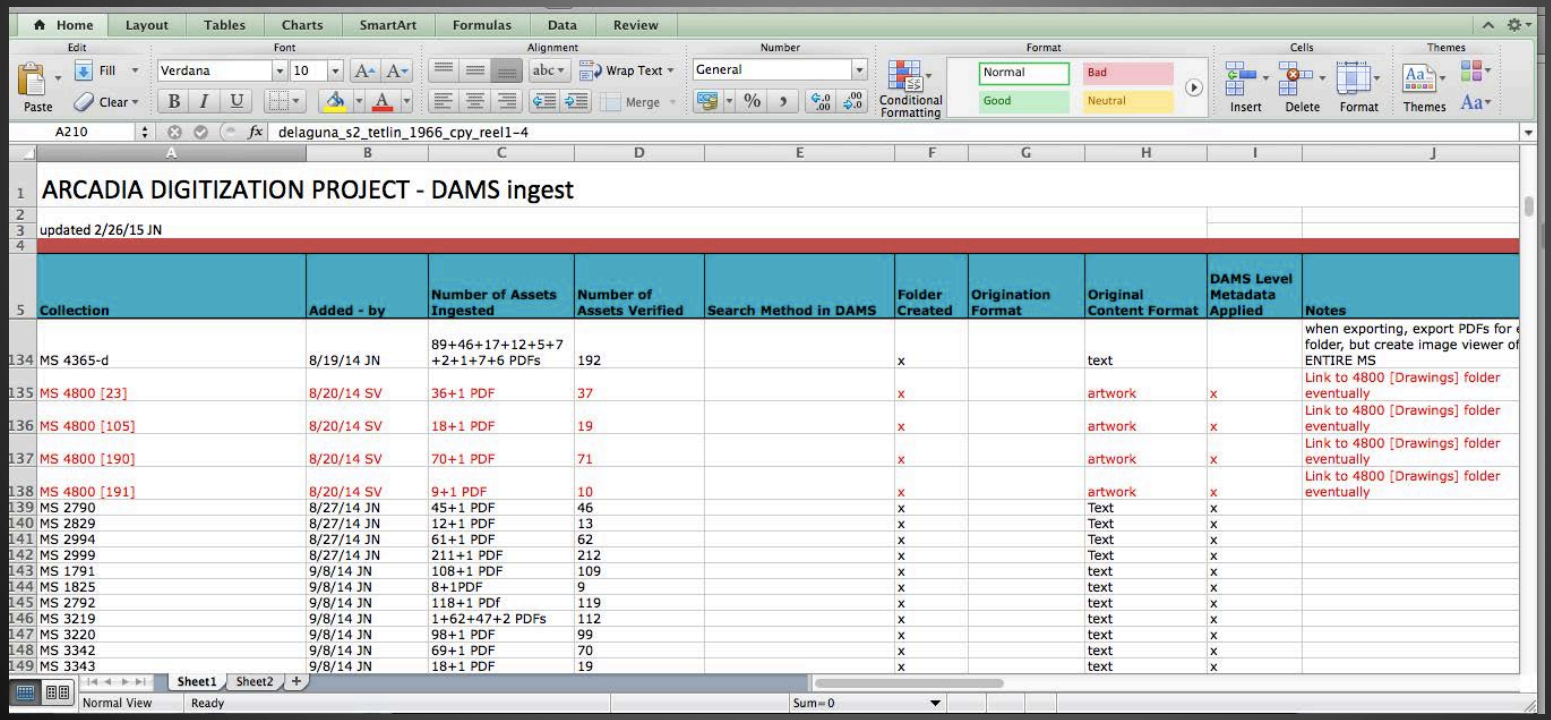
DAMS workflow - 2

- **NEGATIVES:**
- XX_neg folder-> MASTER
- XX_post folder-> SUBFILES
- XX_crop folder-> SUB-SUBFILES
- XX_jpg folder-> do not ingest
- **FILES WITH CROPS AND VERSOS**
- XX.tif file -> MASTER
- XX_crops folder and/ or Versos -> SUBFILES
- XX Verso Crops (if they have been renamed) -> SUB-SUBFILES
- **FILES WITH CROPS AND VERSOS** [Alternative, usually Photo Lots]
- XX.tif file (recto) -> MASTER
- XX.tif file (verso) -> MASTER
- XX crop (recto) -> SUBFILE
- XX crop (verso) -> SUBFILE
- **FILES WITH CROPS AND NO VERSOS**
- XX.tif file -> MASTER
- XX_crops folder -> SUBFILES
- **FILES WITH VERSOS AND NO CROPS**
- XX.tif file -> MASTER
- XX Verso Crops -> SUBFILES

Smithsonian
National Museum of Natural History

DAMS workflow - 3

- Confirm ingest
- Track ingest



ARCADIA DIGITIZATION PROJECT - DAMS ingest

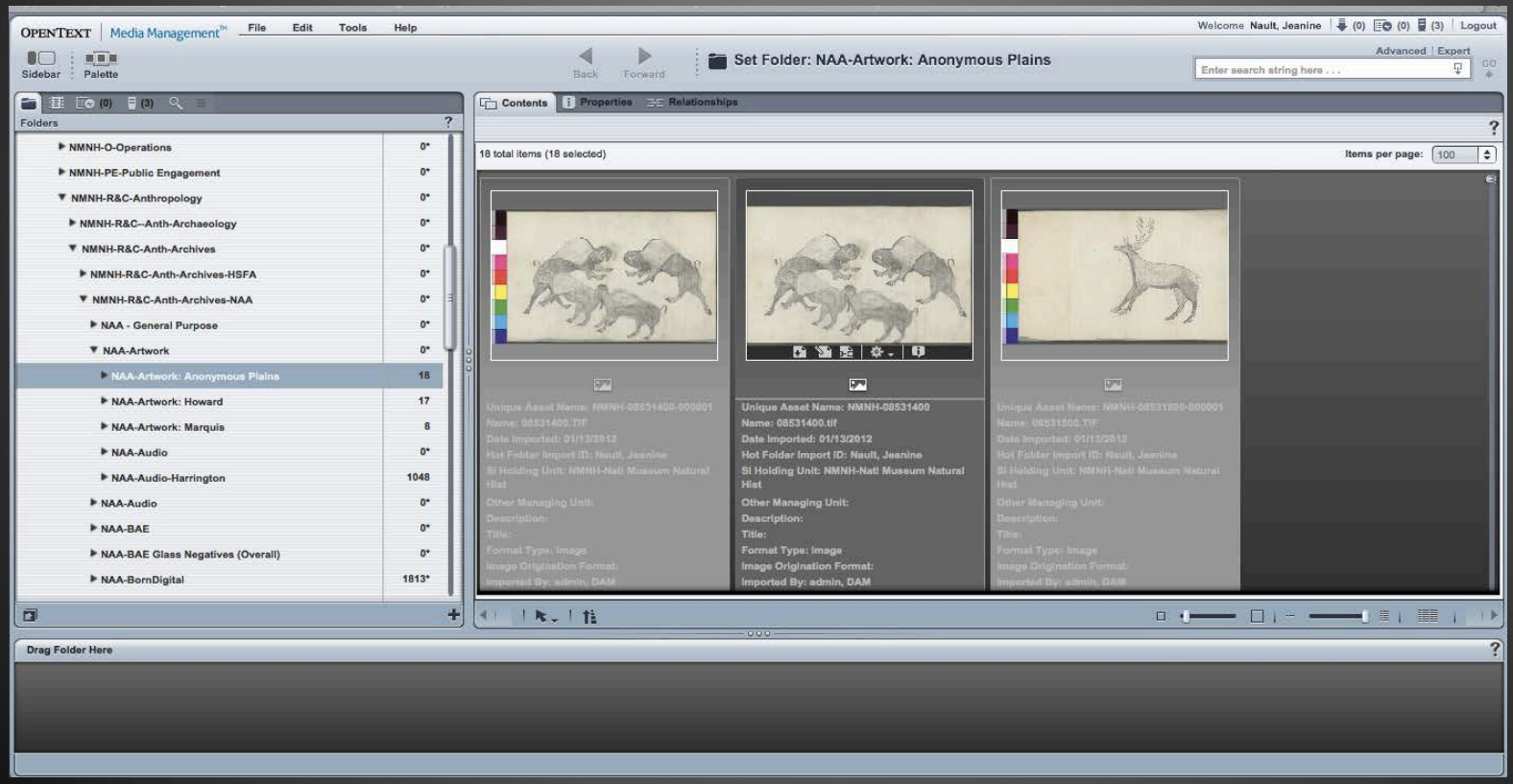
updated 2/26/15 JN

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	Notes
MS 4365-d	8/19/14 JN	89+46+17+12+5+7 +2+1+7+6 PDFs	192		x		text		when exporting, export PDFs for e folder, but create image viewer of ENTIRE MS
MS 4800 [23]	8/20/14 SV	36+1 PDF	37		x		artwork	x	Link to 4800 [Drawings] folder eventually
MS 4800 [105]	8/20/14 SV	18+1 PDF	19		x		artwork	x	Link to 4800 [Drawings] folder eventually
MS 4800 [190]	8/20/14 SV	70+1 PDF	71		x		artwork	x	Link to 4800 [Drawings] folder eventually
MS 4800 [191]	8/20/14 SV	9+1 PDF	10		x		artwork	x	Link to 4800 [Drawings] folder eventually
MS 2790	8/27/14 JN	45+1 PDF	46		x		Text	x	
MS 2829	8/27/14 JN	12+1 PDF	13		x		Text	x	
MS 2994	8/27/14 JN	61+1 PDF	62		x		Text	x	
MS 2999	8/27/14 JN	211+1 PDF	212		x		Text	x	
MS 1791	9/8/14 JN	108+1 PDF	109		x		text	x	
MS 1825	9/8/14 JN	8+1PDF	9		x		text	x	
MS 2792	9/8/14 JN	118+1 PDf	119		x		text	x	
MS 3219	9/8/14 JN	1+62+47+2 PDFs	112		x		text	x	
MS 3220	9/8/14 JN	98+1 PDF	99		x		text	x	
MS 3342	9/8/14 JN	69+1 PDF	70		x		text	x	
MS 3343	9/8/14 JN	18+1 PDF	19		x		text	x	

Smithsonian
National Museum of Natural History

DAMS workflow - 4

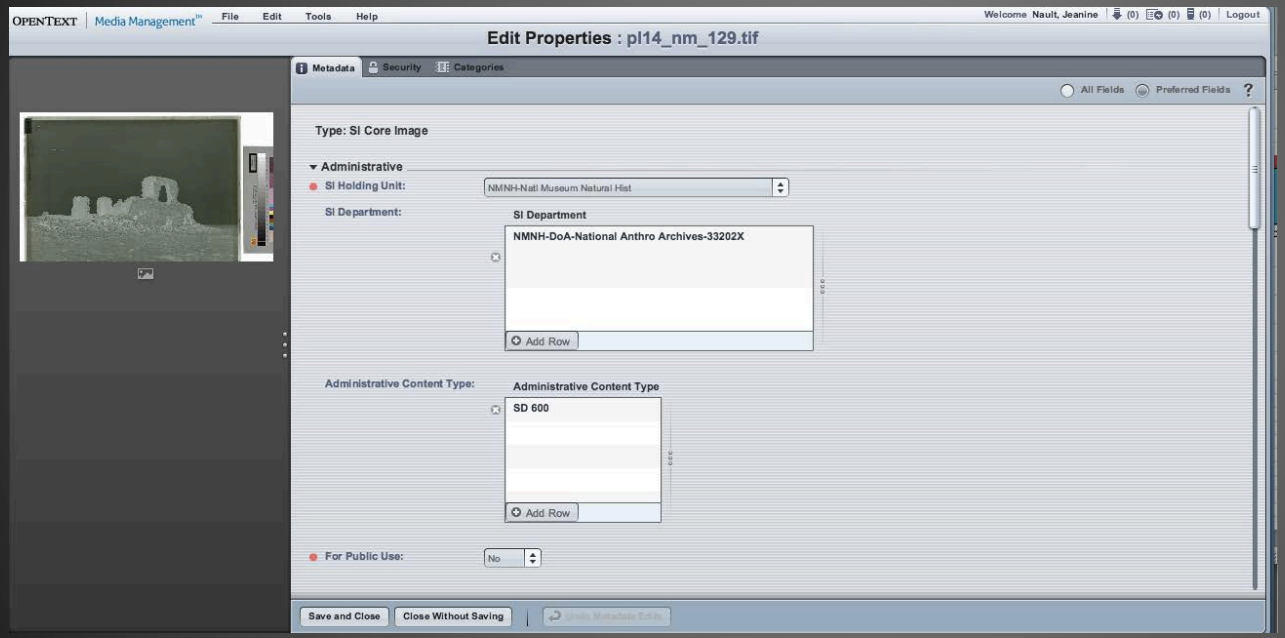
- Search and verify assets



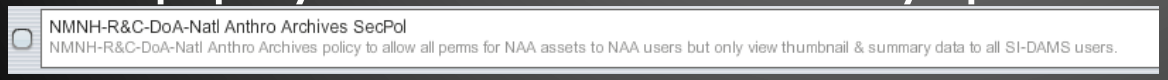
Smithsonian
National Museum of Natural History

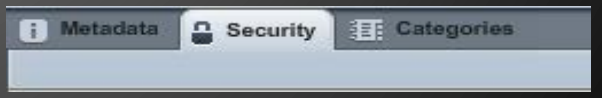
DAMS workflow - 5

- Apply DAMS level metadata



- Apply DAMS security policies

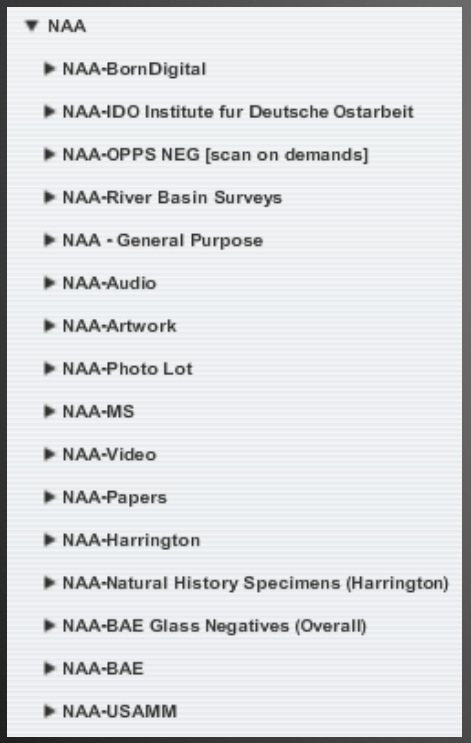


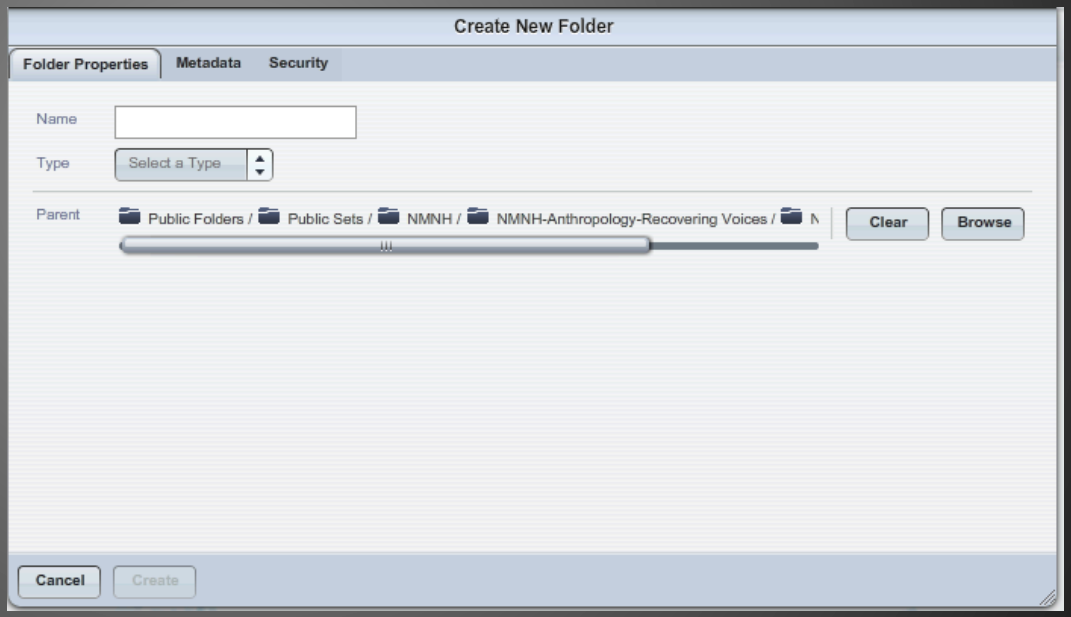


Smithsonian
National Museum of Natural History

DAMS workflow – 6

- Create a Folder with Collection Name and Number



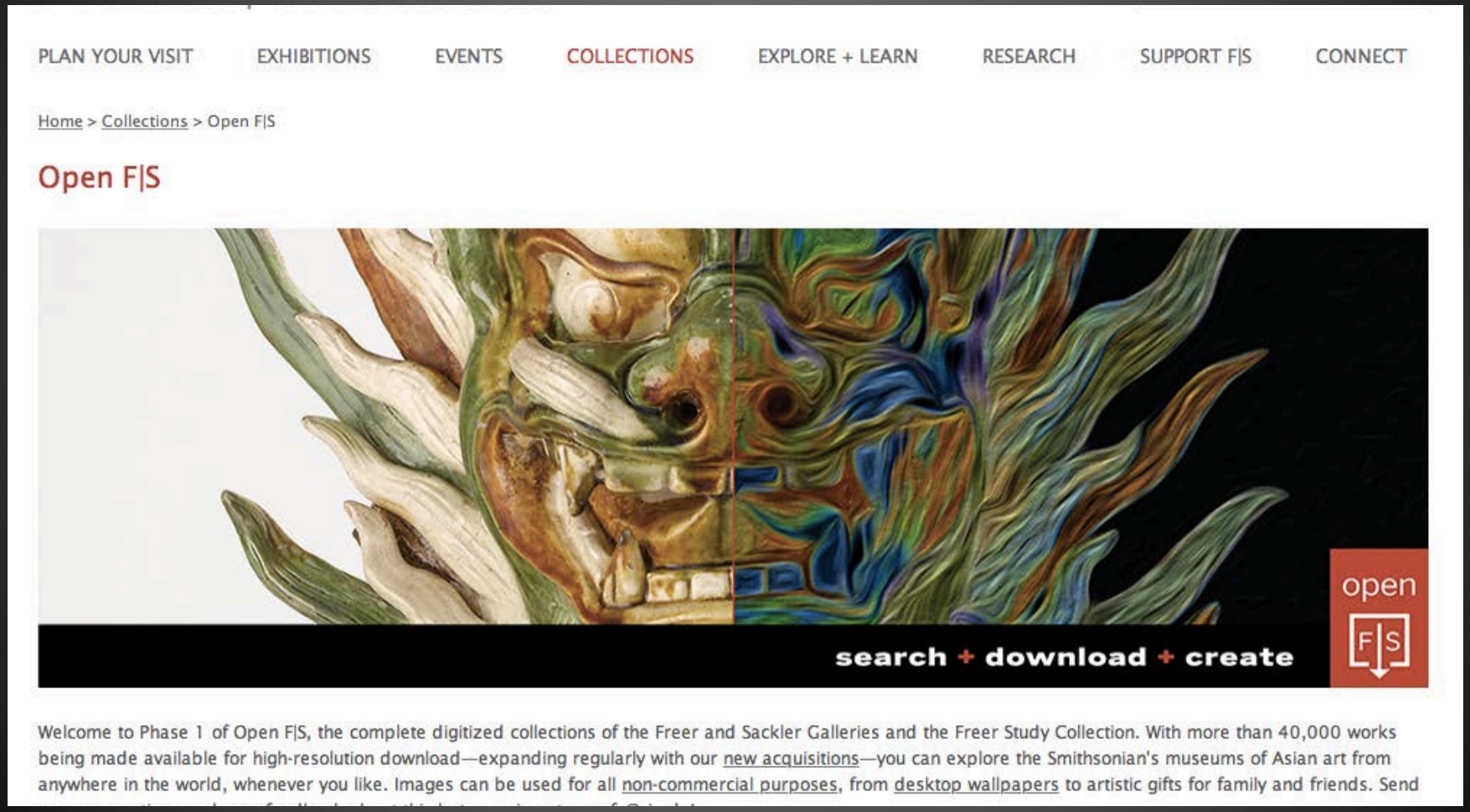


Storage Conclusions

- Data needs to be backed up in multiple locations, migrated, and takes up space
- Decide which storage and/or digital asset management system is right for you
  - infrastructure, support, staff, time, money and NEEDS

Smithsonian
National Museum of Natural History

All About Access



Smithsonian
National Museum of Natural History

Allowing Access

Why?

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

How?

- 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

Smithsonian
National Museum of Natural History

Open Access

[The Getty Open Content Program]

[Digita Vaticana]

Smithsonian
National Museum of Natural History

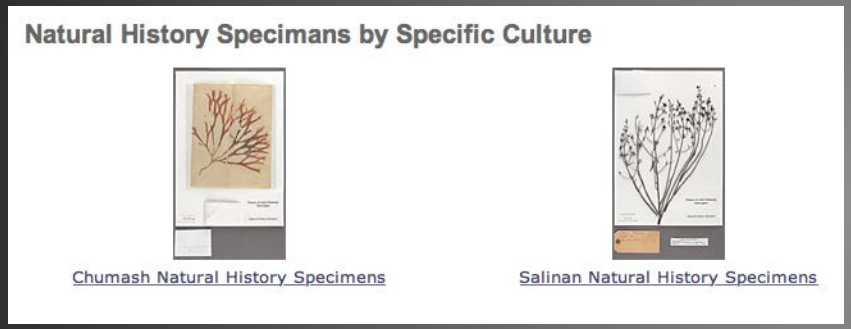
Increasing Access

John P. Harrington web portal

- Digital access to various collections
  - Manuscript and Microfilm
  - Photographs
  - Sound Recordings
  - Botanical specimens

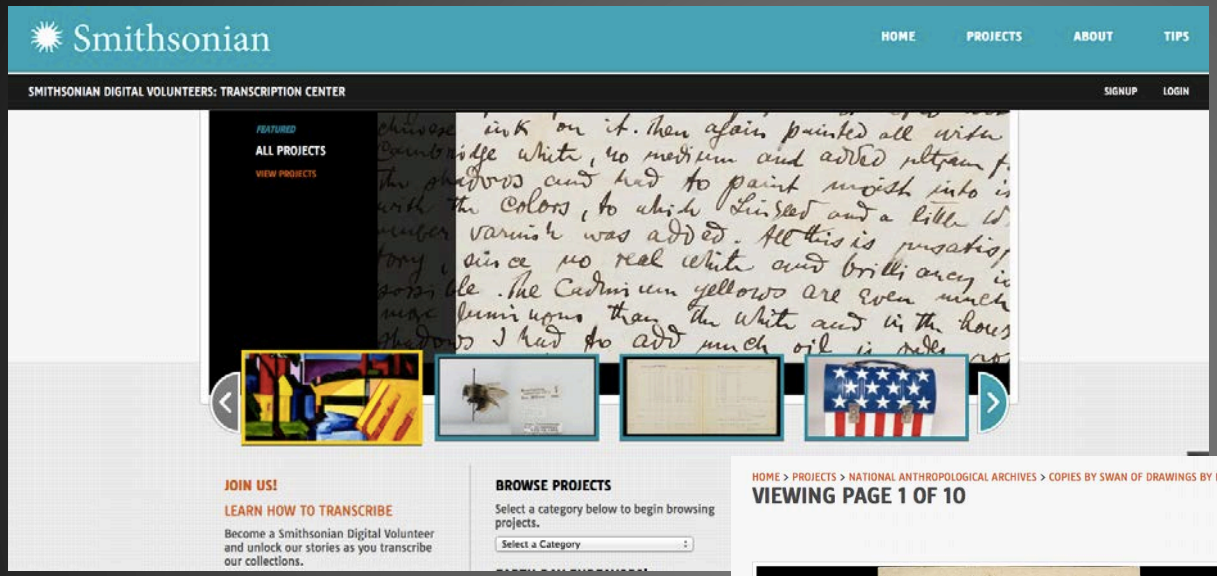
Smithsonian
National Museum of Natural History

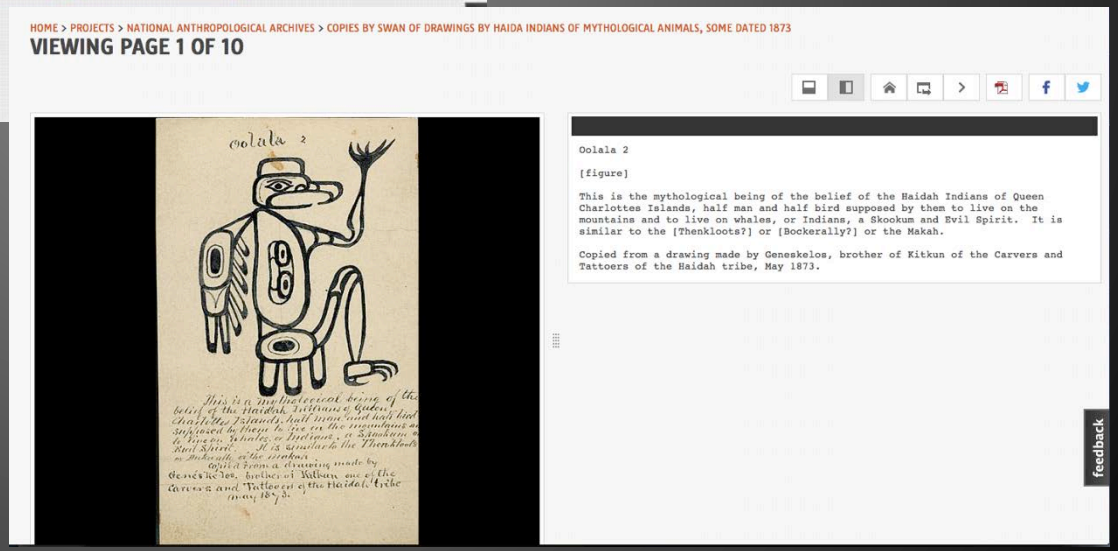
Increasing Access – Harrington web portal



Smithsonian
National Museum of Natural History

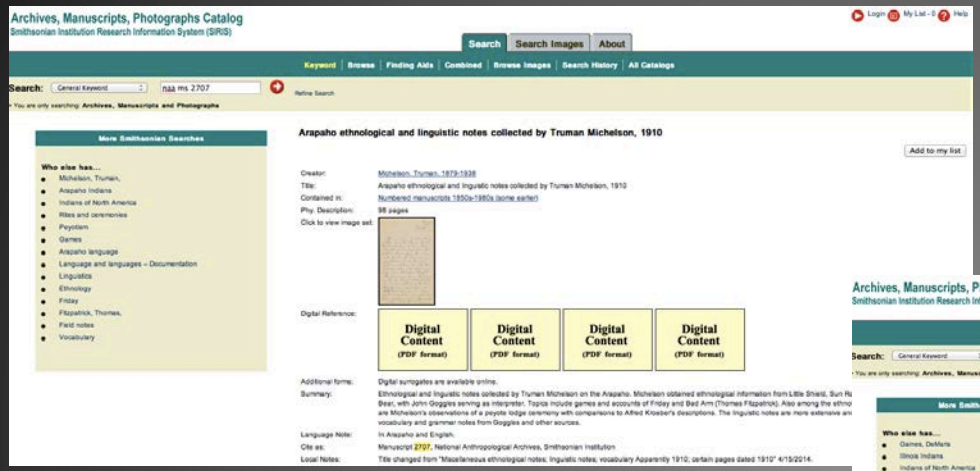
Increasing Access – transcription.si.edu

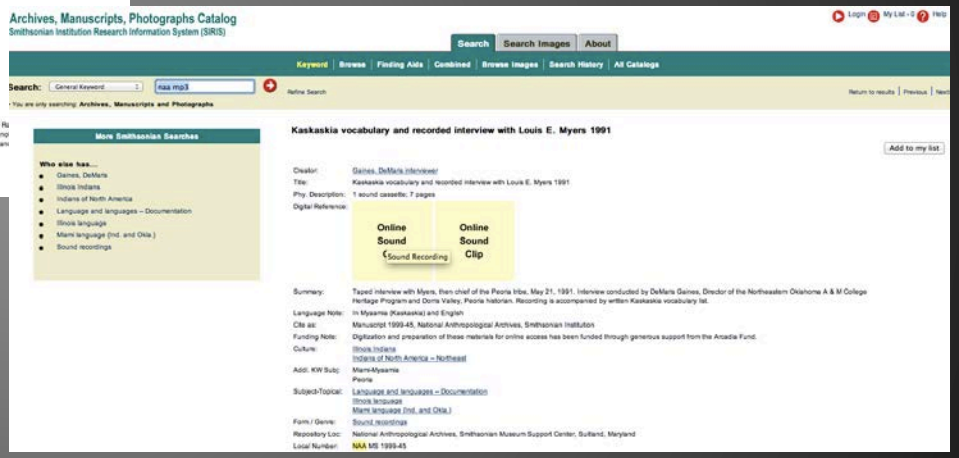




Smithsonian
National Museum of Natural History

Increasing Access – Arcadia project

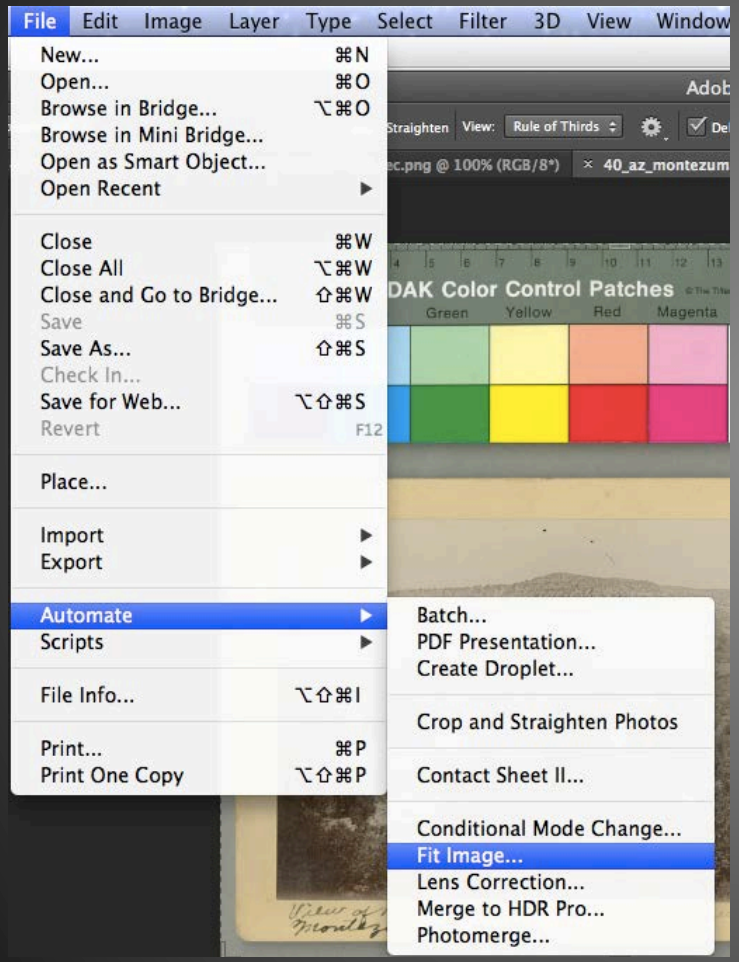




- Freely available, downloadable PDFs of 20,000 pages of endangered language manuscript material and 15,000 pages of associated text related to sound recordings
- Online mp3's of 3500 sound recordings

Smithsonian
National Museum of Natural History

How to create access – derivatives

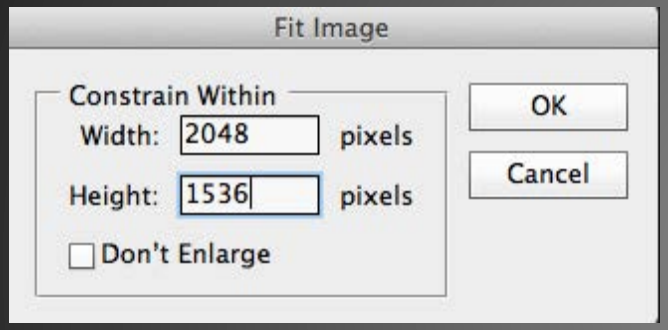


Go to File > Automate > Fit Image

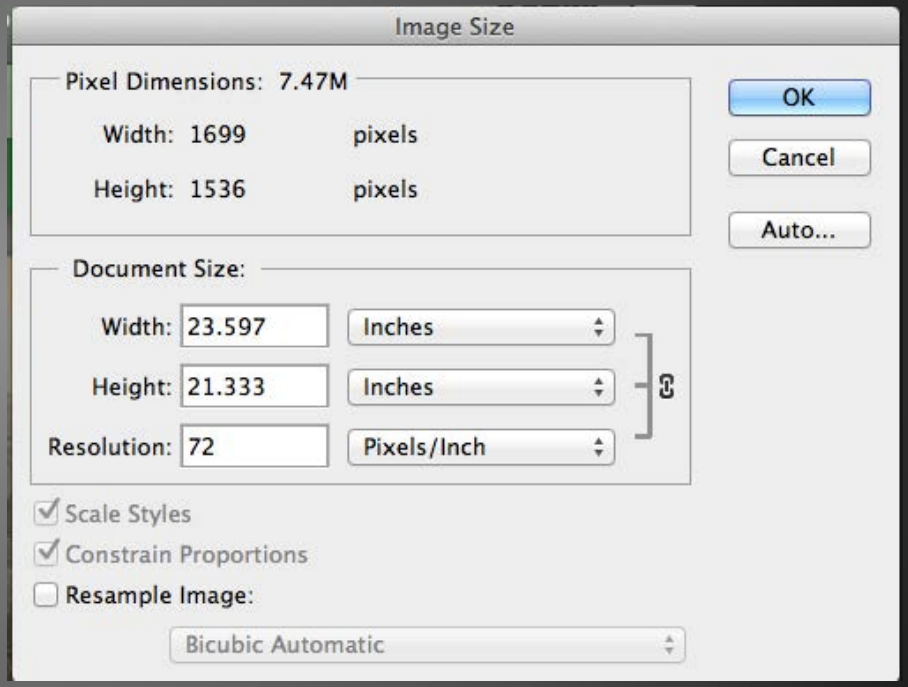
Smithsonian
National Museum of Natural History

How to create access - derivatives

A window will pop up – enter the follow dimensions into the appropriate boxes – click OK



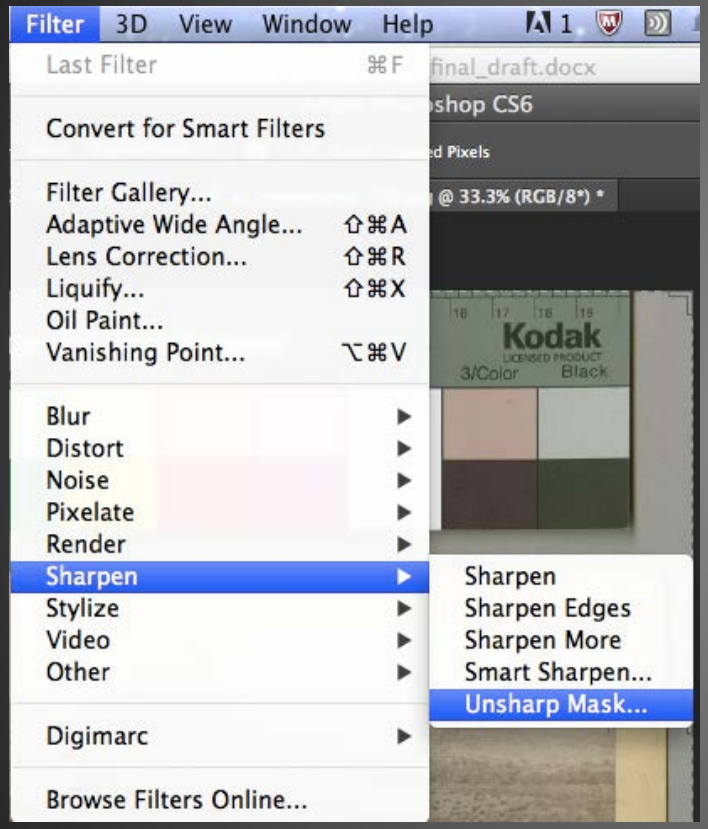
Go to Image > Image Size. An Image Size window will pop up.

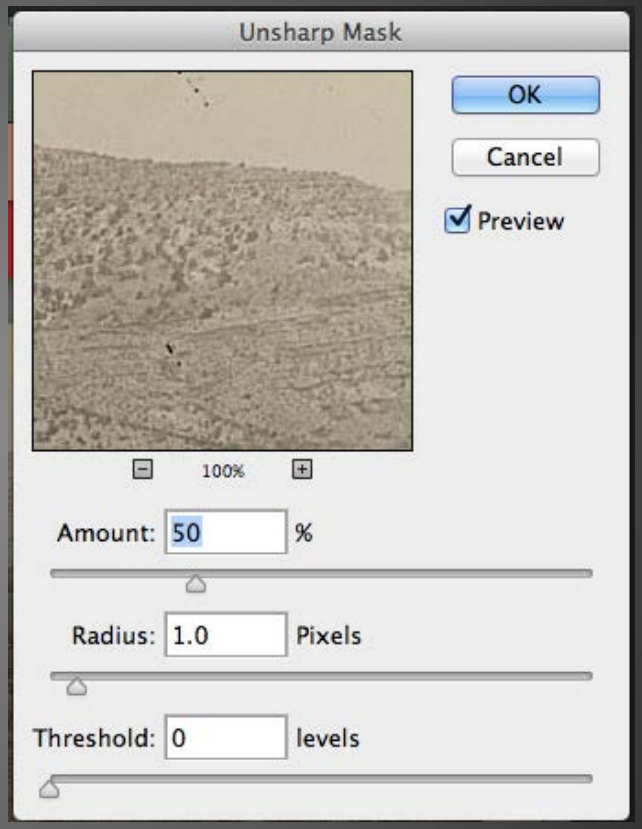


In the resolution field, enter 72. Make sure pixels/inch is selected. Make sure box for "Resample Image" is unchecked. Click ok.

Smithsonian
National Museum of Natural History

How to create access - derivatives



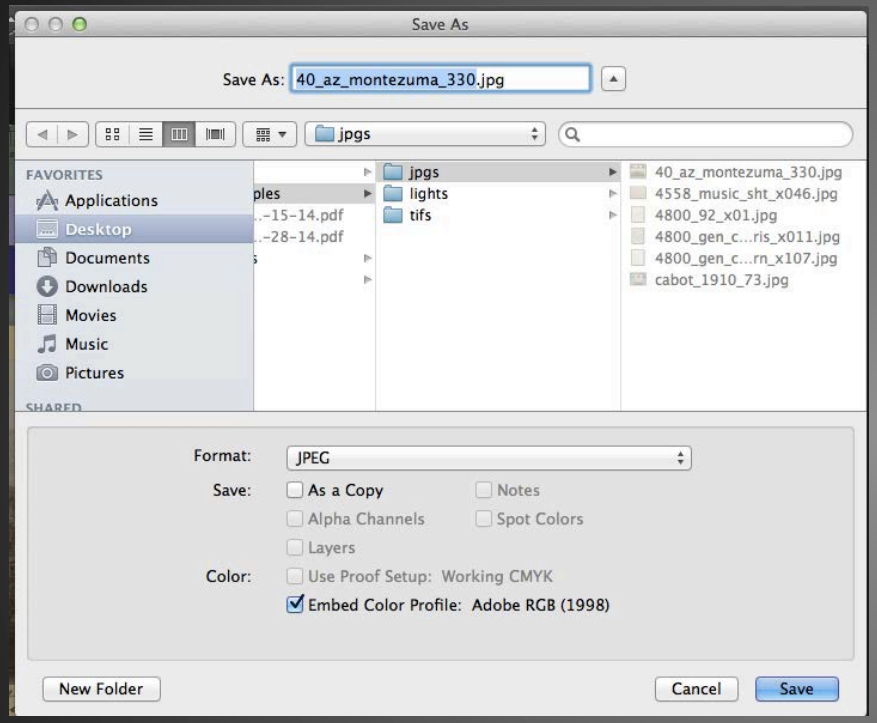


Go to Filter > Sharpen > Unsharp Mask

In the Unsharp Mask window enter the above in each field. Click Ok.

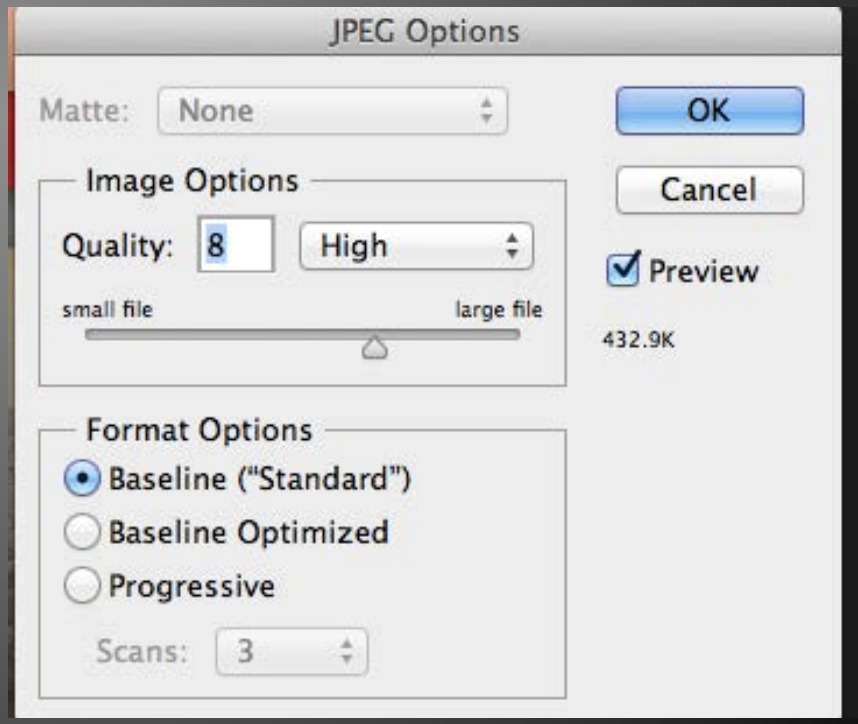
Smithsonian
National Museum of Natural History

How to create access - derivatives



Go to File > Save As. In the Save As window, make sure the format is "jpg," and the box for "Embed Color Profile" is checked. Click Save.

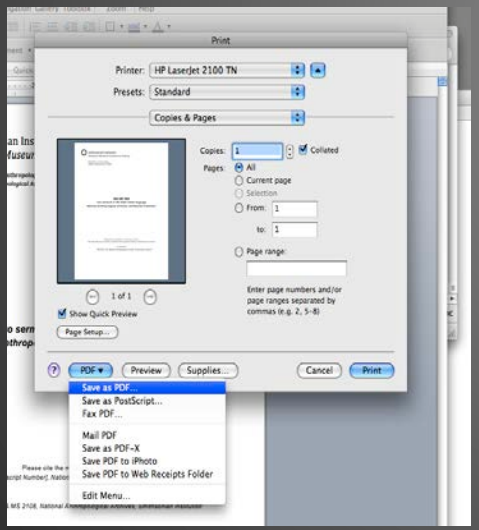
In the JPEG options window, make sure the options are:
Quality: 8
Baseline: Standard
Click Ok. Close window.

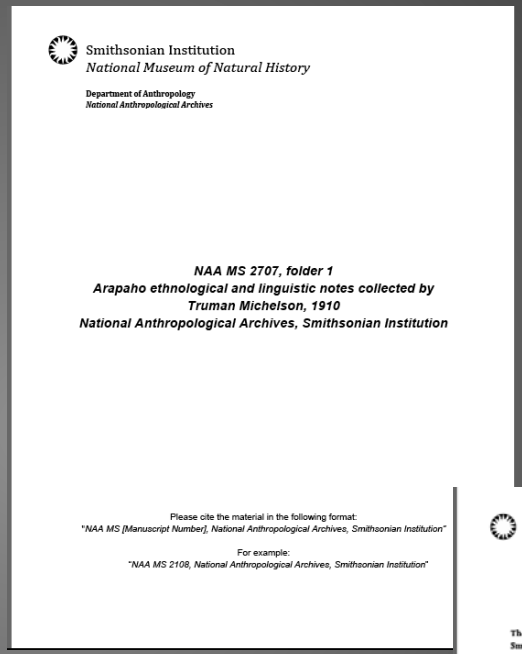


Smithsonian
National Museum of Natural History

How to create access - derivatives

You may want to create a Word file of the designated PDF Cover Sheets. There are both generic cover sheets and project or material specific cover sheets.



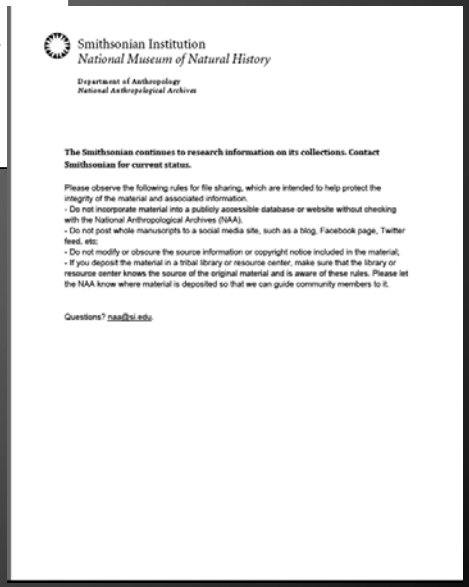

Smithsonian Institution
National Museum of Natural History

Department of Anthropology
National Anthropological Archives

NAA MS 2707, folder 1
Arapaho ethnological and linguistic notes collected by Truman Michelson, 1910
National Anthropological Archives, Smithsonian Institution

Please cite the material in the following format:
"NAA MS [Manuscript Number], National Anthropological Archives, Smithsonian Institution"

For example:
"NAA MS 2108, National Anthropological Archives, Smithsonian Institution"


Smithsonian Institution
National Museum of Natural History

Department of Anthropology
National Anthropological Archives

The Smithsonian continues to research information on its collections. Contact Smithsonian for current status.

Please observe the following rules for file sharing, which are intended to help protect the integrity of the material and associated information.
- Do not incorporate material into a publicly accessible database or website without checking with the National Anthropological Archives (NAA).
- Do not post whole manuscripts to a social media site, such as a blog, Facebook page, Twitter feed, etc.
- Do not modify or obscure the source information or copyright notice included in the material.
- If you deposit the material in a tribal library or resource center, make sure that the library or resource center knows the source of the original material and is aware of these rules. Please let the NAA know where material is deposited so that we can guide community members to it.

Questions? naa@si.edu.

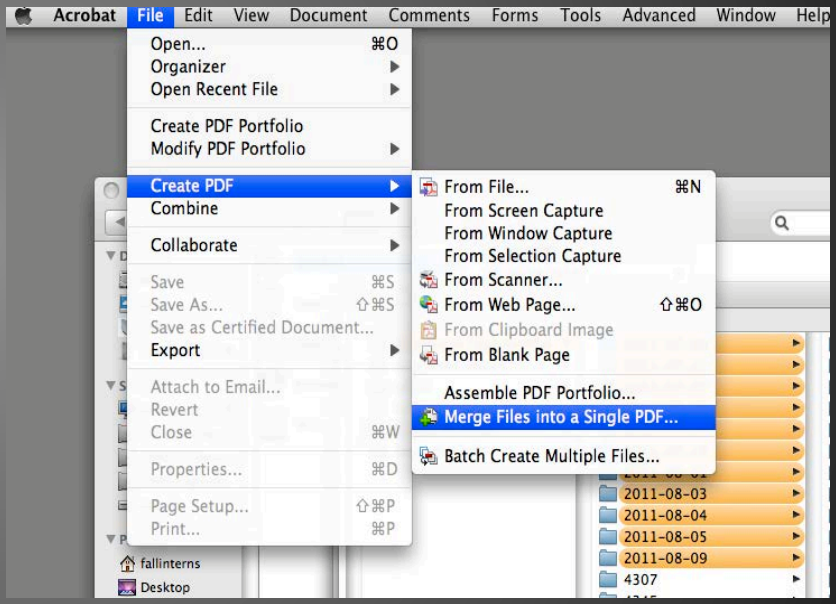
Smithsonian
National Museum of Natural History

How to create access - derivatives

Combining Files and Cover Sheet into PDF file.

Open Adobe Acrobat Professional.

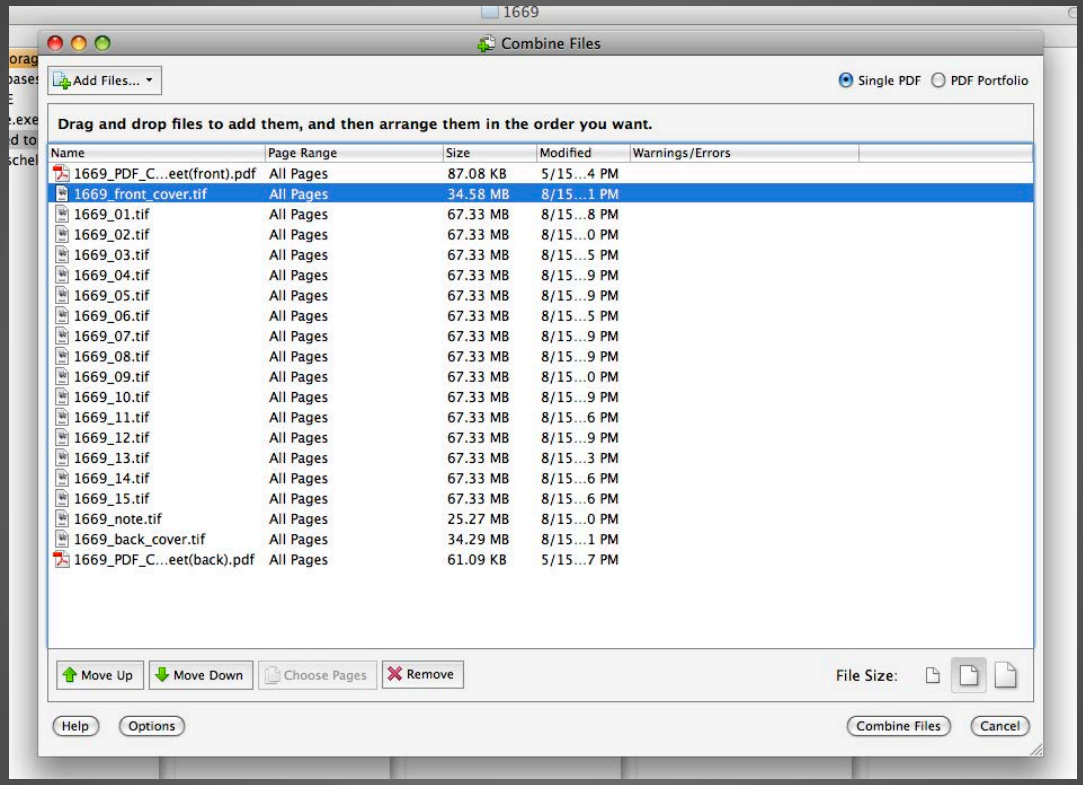
Or you may choose File→Create PDF/or Combine→Merge/ Combine Files into a Single PDF



A Start Menu may pop up. You may choose "Combine Files into PDF"

Smithsonian
National Museum of Natural History

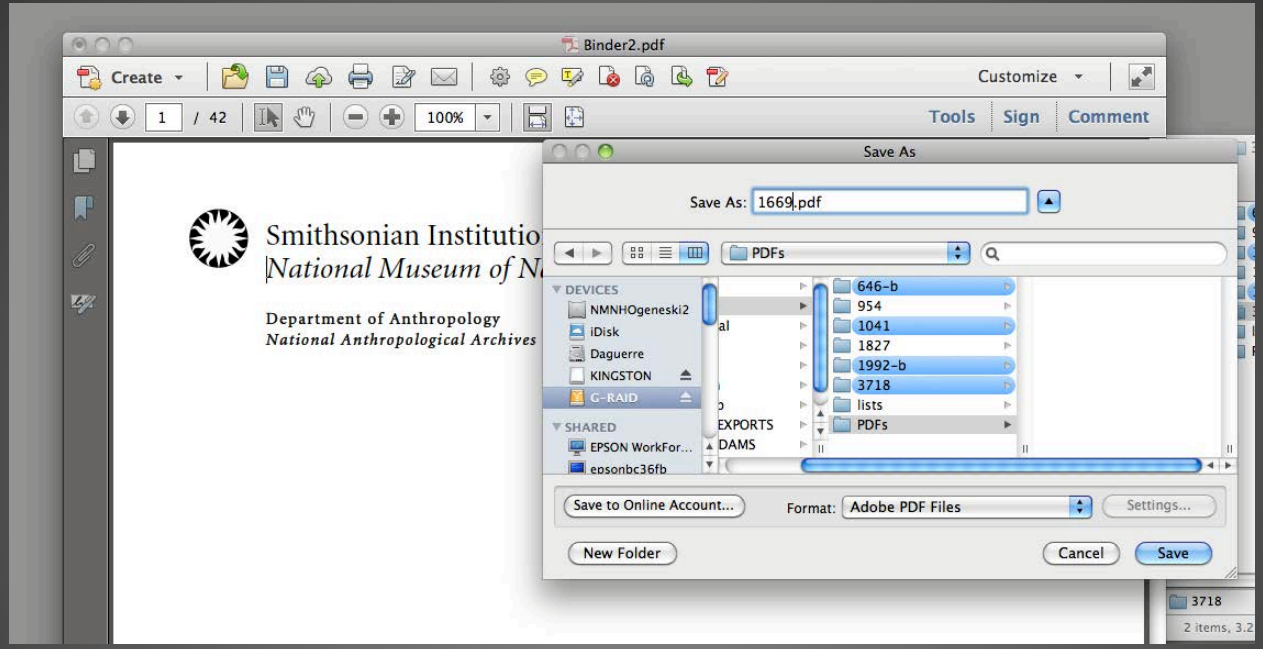
How to create access - derivatives



In the Combine Files window that opens select "Add Files" or drag files and PDF cover sheets into the window. Make sure Single PDF is checked and make sure the file size icon is on the middle setting. Add the individual images to be combined, or a whole folder containing them all.

Smithsonian
National Museum of Natural History

How to create access - derivatives

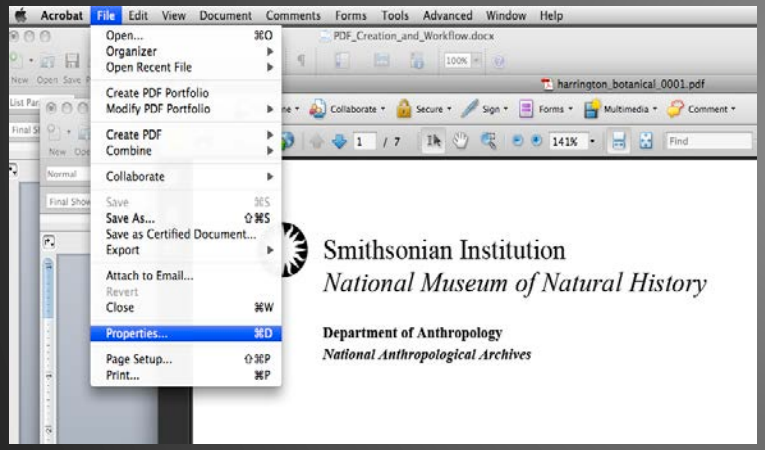


Click Combine Files. Depending on the number of images, it may take a few or several minutes to finish; you can watch the progress at the bottom of the window. Once the PDF is created, it will appear in a new window, starting with the cover sheet. From here, go to File→Save.

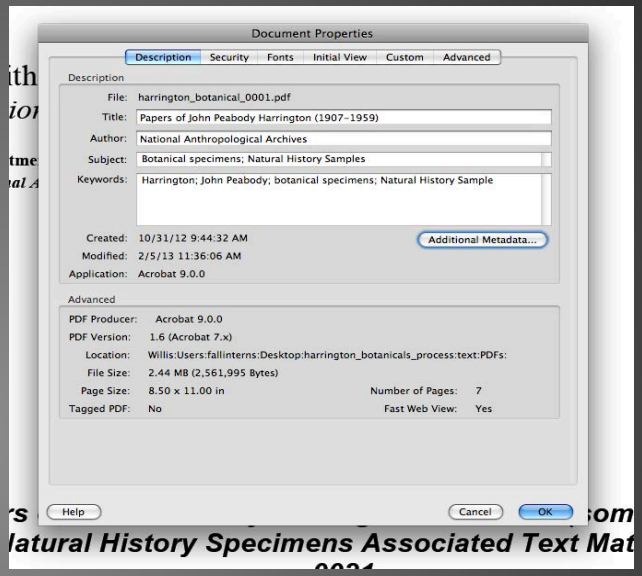
Smithsonian
National Museum of Natural History

How to create access – derivatives

Once the file has been saved, go to File→Properties. A "Document Properties" window will pop up



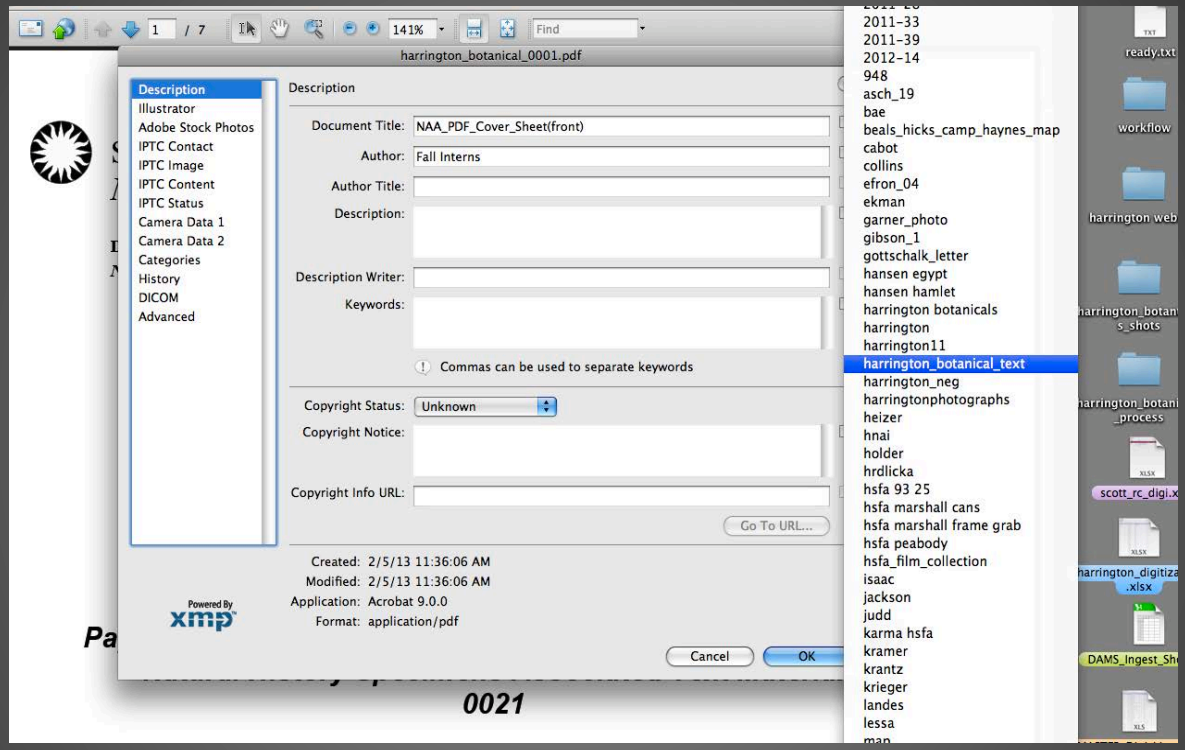
(If the metadata template has been previously saved/exported, it will be available in Acrobat.)



Click "Additional Metadata"

Smithsonian
National Museum of Natural History

How to create access - derivatives



A File Info window will pop up (it will look similar to the one in Photoshop). Click the arrow to the right of "Description." A list of metadata templates will appear. Choose the appropriate template. The fields will populate. Click Ok. Save PDF before closing.

Smithsonian
National Museum of Natural History

ANY QUESTIONS?!

THANK YOU!

naultj@si.edu