



Smithsonian Institution
National Museum of Natural History

Department of Anthropology
National Anthropological Archives

**Bridging the Digital Divide: Basic Guidelines and Best Practices
for Digitization Projects**

**Instructor: Jeanine Nault
Digital Imaging Specialist
National Anthropological Archives
National Museum of Natural History
Smithsonian Institution**

Please cite the material in the following format:
"Courtesy of National Anthropological Archives, Smithsonian Institution"

Digital Imaging Guidelines

National Anthropological Archives Smithsonian Institution

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collaborative effort by federal agencies formed in 2007 to define common guidelines, methods, and practices to digitize historical content in a sustainable manner.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 NAA's imaging guidelines were developed to meet the archives' needs at a particular point in time and reflect its staffing and funding opportunities. Although these standards have served us well, other imaging programs may require different solutions. Our guidelines are therefore not recommendations.

Image Resolution

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. For example, a 5-inch print on an 8-inch mat should be scanned at 1000 dpi ($5 \times 1000 = 5,000$), even though the longest dimension becomes 8,000 pixels ($8 \times 1000 = 8,000$). This exception is made so that the NAA can later reproduce the photograph at 500 dpi.

File Format

Files should be saved in:

Color: 24-bit RGB color TIFF format with an IBM PC Byte Order.

Grayscale: 8-bit Grayscale TIFF format with an IBM PC Byte Order.

Black-and-white photographic negatives should be scanned as RGB and saved as 16 bit grayscale images. "Post" files will be at 8-bit grayscale. There may be exceptions when negatives are saved as RGB files.

ICC Color Profiles

Color: Adobe RGB 1998

Grayscale: Gray Gamma 2.2

In certain cases an ICC-compliant profile may be used and should be documented and included with the project.

Test Targets

Test targets utilized in the NAA imaging lab include:

Kodak Q-13 Color Separation Guide, Kodak #1A Uncalibrated Step Tablet, Gretag Macbeth Color, Checker, Dice Device Characterization Target, Dice Object Level Target, QA-61-ISO-16067-1

Test targets should not be considered when calculating the dimensions of the digital image. A calibration guide should be captured as part of each batch or within each image capture depending on project dictates.

Embedded Metadata

All digital images are embedded with minimal descriptive embedded metadata as described in the document

[*Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images*](#) by EMDaWG (Embedded Metadata Working Group-Smithsonian Institution) April 2010.

Quality Control (QC) and Quality Assurance (QA)

Quality Control (QC) and Quality Assurance (QA) routines are consistent with those outlined in section VIII Quality Management (pp.82-86) of the [Technical Guidelines for Digitizing Cultural Heritage Materials](#), released by FADGI in 2010

File Naming

Please see the document [NAA Digital Filenaming Conventions](#) for more details.

Storage

Upon the completion of each imaging project, images are ingested into the Smithsonian's Digital Asset Management System (DAMS) maintained by the OCIO. All images are ingested with minimal embedded metadata as outlined above, and then a basic DAMS Image Metadata model is applied. Once all images are ingested and verified, they are then available to the Rights Coordinator to apply rights metadata to the assets.

Digital assets are also stored on a local Network Attached Storage (NAS) drive, and an off site tape backup at the Smithsonian's Data Center.

Previous backup of digital assets was on CD and duplicate CD-ROMs. These assets have been migrated to the above backup systems and the CD's have been moved to archival storage.

Project Tracking

A Digitization and Digital Asset Management Policy (SD610) is to be recorded for the project. Each completed digitized collection should also be entered in the NAA's imaging projects database. Each record within the database should indicate:

Accession number and title of collection

Number of original items, number of digital images Items not scanned (and the reason; e.g., broken, fragile, oversize, etc.)

Whether item-and collection-level records exist in the Smithsonian's online public access catalog (SIRIS), as well as which item-level records are missing

Notes on imaging, such as as exceptions to file-naming conventions

Imaging hardware used

Names of imaging specialist and the specialist who proofed the images, with dates for each.

Backup locations, and other project related notation.

Other available derivatives (i.e. PDF)

A Digital Surrogate Sheet noting the date and filenaming and/ or items digitized should be filled out in pencil and included in the analog collection folder so to ensure when viewing the analog collection digitization of particular collections can be easily identified. If the entire box is digitized only one sheet is required at the beginning of the box.

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-

descriptive (such as a sequential numerical string). Meaningful file names contain metadata that is self-referencing; non-descriptive file names are associated with metadata stored elsewhere that serves to identify the file. In general, smaller-scale projects may design descriptive file names that facilitate browsing and retrieval; large-scale projects may use machine-generated names and rely on the database for sophisticated searching and retrieval of associated metadata.

A file naming system based on non-descriptive, non-mnemonic, unique identifiers usually requires a limited amount of metadata to be embedded within the file header, as well as an external database which would include descriptive, technical, and administrative metadata from the source object and the related digital files.

One advantage of a non-descriptive file naming convention is that it eliminates non-unique and changeable descriptive data and provides each file with a non-repeating and sustainable identifier in a form that is not content-dependent. This allows much greater flexibility for automated data processing and migration into future systems. Other benefits of a non-descriptive file naming convention include the ability to compensate for multiple object identifiers and the flexibility of an external database, which can accommodate structural metadata including parts and related objects, as well as avoiding any pitfalls associated with legacy file identifiers.

In general, we recommend that file names

- Are unique - no other digital resource should duplicate or share the same identifier as another resource. In a meaningful file-naming scheme, names of related resources may be similar, but will often have different characters, prefixes, or suffixes appended to delineate certain characteristics of the file. An attempt to streamline multiple versions and/or copies should be made.
- Are consistently structured - file names should follow a consistent pattern and contain consistent information to aid in identification of the file as well as management of all digital resources in a similar manner. All files created in digitization projects should contain this same information in the same defined sequence.
- Are well-defined - a well-defined rationale for how/why files are named assists with standardization and consistency in naming and will ease in identification of files during the digitization process and long afterwards. An approach to file naming should be formalized for digitization projects and integrated into systems that manage digital resources.
- Are persistent – files should be named in a manner that has relevance over time and is not tied to any one process or system. Information represented in a file name should not refer to anything that might change over time. The concept of persistent identifiers is often linked to file names in an online environment that remain persistent and relevant across location changes or changes in protocols to access the file.

- Observant of any technical restrictions – file names should be compliant with any character restrictions (such as the use of special characters, spaces, or periods in the name, except in front of the file extension), as well as with any limitations on character length. Ideally, file names should not contain too many characters. Most current operating systems can handle long file names, although some applications will truncate file names in order to open the file, and certain types of networking protocols and file directory systems will shorten file names during transfer. Best practice is to limit character length to no more than 32 characters per file name.
- We recommend using a period followed by a three-character file extension at the end of all file names for identification of data format (for example, .tif, .jpg, .gif, .pdf, .wav, .mpg, etc.) A file format extension must always be present.
- Take into account the maximum number of items to be scanned and reflect that in the number of digits used (if following a numerical scheme).
- Use leading 0's to facilitate sorting in numerical order (if following a numerical scheme).
- Do not use an overly complex or lengthy naming scheme that is susceptible to human error during manual input.
- Use lowercase characters and file extensions.
- Record metadata embedded in file names (such as scan date, page number, etc.) in another location in addition to the file name. This provides a safety net for moving files across systems in the future, in the event that they must be renamed.
- In particular, sequencing information and major structural divisions of multi-part objects should be explicitly recorded in the structural metadata and not only embedded in filenames.
- Although it is not recommended to embed too much information into the file name, a certain amount of information can serve as minimal descriptive metadata for the file, as an economical alternative to the provision of richer data elsewhere.
- Alternatively, if meaning is judged to be temporal, it may be more practical to use a simple numbering system. An intellectually meaningful name will then have to be correlated with the digital resource in an external database.

Directory Structure

Regardless of file name, files will likely be organized in some kind of file directory system that will link to metadata stored elsewhere in a database. Master files might be stored separately from derivative files, or directories may have their own organization independent of the image files, such as folders arranged by date or collection identifier, or they may replicate the physical or logical organization of the originals being scanned.

The files themselves can also be organized solely by directory structure

and folders rather than embedding meaning in the file name. This approach generally works well for multi-page items. Images are uniquely identified and aggregated at the level of the logical object (i.e., a book, a chapter, an issue, etc.), which requires that the folders or directories be named descriptively. The file names of the individual images themselves are unique only within each directory, but not across directories. For example, book 0001 contains image files 001.tif, 002.tif, 003.tif, etc. Book 0002 contains image files 001.tif, 002.tif, and 003.tif. The danger with this approach is that if individual images are separated from their parent directory, they will be indistinguishable from images in a different directory.

Versioning

For various reasons, a single scanned object may have multiple but differing versions associated with it (for example, the same image prepped for different output intents, versions with additional edits, layers, or alpha channels that are worth saving, versions scanned on different scanners, scanned from different original media, scanned at different times by different scanner operators, etc.). Ideally, the description and intent of different versions should be reflected in the metadata; but if the naming convention is consistent, distinguishing versions in the file name will allow for quick identification of a particular image. Like derivative files, this usually implies the application of a qualifier to part of the file name. The reason to use qualifiers rather than entirely new names is to keep all versions associated with a logical object under the same identifier. An approach to naming versions should be well thought out; adding 001, 002, etc. to the base file name to indicate different versions is an option; however, if 001 and 002 already denote page numbers, a different approach will be required.

Naming Derivative Files

The file naming system should also take into account the creation of derivative image files made from the master files. In general, derivative file names are inherited from the masters, usually with a qualifier added on to distinguish the role of the derivative from other files (i.e., “pr” for printing version, “t” for thumbnail, etc.) Derived files usually imply a change in image dimensions, image resolution, and/or file format from the master. Derivative file names do not have to be descriptive as long as they can be linked back to the master file.

For derivative files intended primarily for Web display, one consideration for naming is that images may need to be cited by users in order to retrieve other higher-quality versions. If so, the derivative file name should contain enough descriptive or numerical meaning to allow for easy retrieval of the original or other digital versions.

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

I. Introduction

This document defines the minimum proposed descriptive embedded metadata¹ for digital images² at the Smithsonian Institution. It also provides basic guidelines for the use of embedded metadata in imaging workflows across the Institution. It represents a consensus of practitioners across various units about how to better manage the data that is populated into our digital images. Its intention is not to dictate practice, but rather to educate and provide guidance for those working with digital image collections.

Though we acknowledge that this information continues to evolve, we also recognize the Institution's desire for clarity about how to identify and use digital materials at this point in time. Further, a commonly understood and implemented metadata terminology will assist in the implementation of the goals and objectives of the SI Strategic Plan.

The group recommends that metadata management criteria be reviewed every two years to stay current with best practices. At the time of this document's publication Adobe had released the latest version of their Creative Suite (CS5), which includes recent updates to the IPTC schema. The working group determined that these changes would not be reflected in the document until CS5 is more widely adopted and evaluated.

This document identifies the following:

1. Required core set of embedded metadata
2. Suggested set of embedded metadata
3. Recommendations of data value types not to embed in digital images
4. Recommendations when working with vendors who will be embedding metadata
5. Appendices:
 - a. IPTC Element Narratives
 - b. Examples of IPTC fields mapped to various imaging programs

The group was established in April 2009 in response to a growing request from staff who work daily with digital assets to clarify practices of embedding metadata in images. The group's goal was to establish a core set of minimal embedded metadata fields and to define the type of data values for those fields which could both provide better online access of Smithsonian images and, ensure preservation of these images in the future. One of the first steps of the group was to survey members of the group as to their current practice, if any. Once compiled, results were compared and key embedded metadata fields were identified. Disparities were found in the metadata fields used by various units, which made sharing common search catalogs difficult. This prompted several discussions as to the types of data values that might better be contained in those fields, and that would serve a majority of units. In an attempt to accommodate the different, various needs of individual units when populating these fields, it was

¹ Technical metadata is outside the scope of this document.

² Digital images includes still image files and born digital files.

determined that an additional suggested set should also be established. The suggested set also contains sample values, provided by different units in the group, for populating fields. During these discussions, the group also advised establishing recommendations for data value types which should not be embedded. They also elaborated on recommendations for embedded metadata when working with scanning or imaging vendors. Finally, the group established an appendix of IPTC fields mapped to various imaging programs. The intention is to guide practitioners when populating fields in various metadata editing utilities and software versions so as to ensure proper mapping.

II. Background

Metadata is a crucial, yet largely invisible set of information about the characteristics of any item, such as who collected it, when it was collected, and in what format. In digital imaging, there are two types of embedded metadata (information stored within the file itself): technical and descriptive. Technical metadata, which may consist of an equipment name, manufacturer, and capture date, is usually embedded automatically into the file by the equipment software that produces the file. Descriptive metadata requires manual input of additional information about the image by a library, museum, archive or information specialist; this input process may be automated to a certain degree, but may still require manual data entry of information that is unique to the image.

The process of producing metadata about images is not new. We all have done this by writing brief description on the front or back of a photograph. In today's digital world, this process is just updated so that we can provide more and more descriptive information which can travel with the file no matter where it goes. Why is this important? Adding descriptive metadata to a digital image allows us to take advantage of existing technologies that can read and extract that metadata, allowing others to search for our images. It is also helpful when an image comes back to the Institution, from a researcher or general public, in determining its original source and location. The more standardized and *useful* information we put into the metadata, the more searchable these images. As the Smithsonian increasingly digitizes and shares its images, the group strongly suggests that the Institution adopts the standards of best practices for embedded minimal metadata.

A minimal amount of metadata within digital assets takes on critical importance when it comes to the Institution's enterprise Digital Asset Management System (DAMS). The SI DAMS provides for an Institution-wide application used for storing, searching and retrieving digital assets. Without the requisite metadata to accompany each digital asset, the DAMS would be reduced to an unorganized storage system filled with millions of files but no efficient way to search on or to retrieve assets. The SI DAMS' extensive Core Metadata Model (see: https://sp.si.edu/col/SI-DAM/Metadata%20Related/0-SI_DAM_CoreModel_Elements-01-19-2010.ppt) is the basis for the DAMS functionality.

Realizing that some of the images created at SI are:

- Already contained within the SI DAMS;
- Currently residing outside the SI DAMS but may eventually be ingested into the SI DAMS;

the group tried to keep in mind how data values would map to fields within the current SI DAMS Core Metadata Model.

The group also tried to keep in mind digitization workflows, productivity and availability of software across the Institution to facilitate practical implementation. We realize that a variety of people--photographers, digital imaging specialists, web practitioners, scientists, interns, vendors, etc.--may be creating digital images. In many cases, people will be working with various databases and Collection Information systems (CIS). Most likely, embedded metadata will be inserted by the asset creator (i.e. photographers, digital imaging specialists, interns). Often in the creation of the digital asset, embedded metadata can be batch processed. Batch processing aids in a more efficient workflow and productivity. Be mindful that data which is unique to a single asset should not be included as part of batch processing. If batch processing is not tenable, then not embedding metadata may be the most efficient option.

III. Recommendations

The group recommends a *minimal core set* of embedded metadata. After surveying the group as to their practice, it became apparent that many units across the institution were using a version of the International Press Telecommunications Council (IPTC) format.³ Because the core set of elements may not serve the entire Institution community; the group provides a recommended suggested set of embedded metadata as an extension to the core set of fields. Depending on the image file format used, IPTC data can be stored within the file itself in a variety of ways. In recent versions of many image viewing and processing programs, embedded IPTC data is increasingly being saved in XMP format⁴. The recommendations in this document do not indicate a preference on which way the IPTC data is stored within the file.

The tables below contain the element name, definition of that element, sample data values, notes, and character field limitations. It should also be noted that when considering digitization, filenaming conventions should be established beforehand and contained in your digital imaging specifications. When repurposing the digital surrogates, i.e. Web use, image fulfillment, we recommend *against* stripping out the embedded metadata for various reasons stated above.

The group has also noted that SAO has its own metadata model, called AVM and is available at the following link:http://www.virtualastronomy.org/avm_1.1_final_draft.pdf

³ Go to www.iptc.org for detailed information regarding IPTC

⁴ Go to www.adobe.com/products/xmp/ for detailed information regarding XMP, and to www.adobe.com/devnet/xmp/ for the XMP specification documents.

Required: Core Set of Embedded Metadata

Element Name	Definition	Sample Data Value	Notes	Character Limit*
Document Title	File number, Accession Number, Catalog Number, Digital File Name, Negative Number, Unique Identifier root level etc.	1) LB016021-a 2) 08596201 3) gn_03644 4) landes_photo_arizona_16 5) 23456.000 6) 123457.000;1234568.000 * 7) P00001 8) N00001	5) example of a catalog number 6) example of multiple objects in one image 7) example of a catalog number 8) example of catalog number 9) example of catalog number	*IPTC fields have character limits depending on the application utilizing the standard and not all data may be displayed. With some applications data may be truncated at a given character limit. 64

		9) T00001 10)20100121_01a_csf_ps_001.tif 11)SFF2009_Strauss_6-24_0004.dng	10) example of file name representing coverage of event at NMAI	
Copyright Notice	Copyright Notice	1) The Smithsonian continues to research information on its collections. Contact Smithsonian for current status. 2) This image is in the public domain. 3) Copyright National Anthropological Archives, Smithsonian Institution 4) Copyright William M. Groethe	Smithsonian staff should provide accurate copyright information particularly if the copyright status is known. The following default statement should only be used if the unit does not know the copyright status of the work. 'The Smithsonian continues to research information on its collections. Contact Smithsonian for current status' ⁵ .	128
Source	Name and Abbreviation of SI owning unit, Smithsonian Institution	1) NMAI-Natl. Museum of the American Indian, Smithsonian Institution 2) NAA- Natl. Anthropological Archives, Smithsonian Institution		32
Creator (*Note: Unit makes decision- document)	Creator of digital object	1) Smithsonian Institution Libraries	*IPTC Creator Job Title field can be used to define the role of the creator.	32

⁵ The committee recognizes that copyright data may change however it is so important that it is better to put in data if known.

<p>within unit how they reached this decision).</p> <p>In the DAMS the IPTC Creator field is mapped to the Asset Creator field, which is the creator of the digital object.</p>	<p>or</p> <p>Creator of original object</p>	<p>2) Department of Anthropology</p> <p>3) National Anthropological Archives</p> <p>4) Photographer Name</p> <p>5) Cynthia Frankenburg</p> <p>6) William Greene</p> <p>7) Woody Guthrie</p>	<p>4) <i>*If author is not known then default to department name (refrain from using acronyms)</i></p> <p>5) <i>*if Creator=name then Creators Title field is populated. Job Title=Photographer</i></p> <p>6) <i>*if Creator=name then Creators Title field is populated. Job Title=Scanner</i></p> <p>7) <i>*Creator original object</i></p>	
---	---	---	---	--

Not Required: Suggested Set of Embedded Metadata

Element Name	Definition	Sample Value	Notes	Character Limit
Date <i>In the DAMS IPTC Date is mapped to Asset Creation Date, which is the creation date of the digital object.</i>	Date of Object <i>or</i> Date of Creation of Digital	1) 07/23/1967 2) 02/14/2009 3) 01/01/2001	Date structure for IPTC is MM/DD/YYYY. Description field can be used if date range or other date structure is being used. If year is only known, then default to the first day of the first month of the year. 01/01/2001 would be the value for sometime during the year 2001.	8
Description	Free narrative text	1) 123457.000 (right); 123458.000 (left) 2) Cultural Resources Center 2007 Powwow Open House, CRC Open House, CRC Exterior, Chief Joseph, Nez Perce 3) scan from 4x5 CT, slide or B&W negative 4) media of original art (oil on canvas, watercolor, etc.)	1)* <i>see above under Title example. This field is used to locate individual objects within an image that contains multiple objects.</i>	2000
Keywords	Free text field but should be used to store a list of standard term(s) separated by a common delimiter such as semicolon.	1) Lighting Archive; Electrification; Lighting; Lighting Fixtures; Architectural History; History of Architecture 2) Object; Publication; Our Lives 3) Alice Fletcher; Francis La Flesche; 4558	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 <i>consistent</i> with in your unit 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	64

		<p>4) Viento de Agua; plena; bomba</p> <p>5) rugby; Wales; sports</p> <p>6) name of exhibition object is photographed for inclusion in</p>	<p>for consideration.)</p> <p>2) * 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.</p>	
Credit/Provider	<p>What you would like to accompany the image in a publication. Ex:</p> <p>Image Number, SI owning Unit, Smithsonian Institution</p>	<p>1) Image Number, NAA, Smithsonian Institution</p> <p>2) Ken Rahaim, Smithsonian Institution</p> <p>3)Photographer's name, museum (i.e. National Portrait Gallery)</p> <p>** Ernest L. Spybuck (Absentee Shawnee, 1883–1949), <i>Procession before War Dance</i>, ca. 1910. Watercolor on paperboard, 42.2 x 63.9 cm. Oklahoma. Photo by David Heald. 2/5735</p>		32
Job Identifier	Instructions or unit id for	MSC07-04608		32

	a job			
Headline	(Formally called Caption) A descriptive title or a caption.	Dr. J.E. Tallmage	This field was formerly known as "Caption". If you are using older software you may still have Caption and not have "Headline"	256

Controlled Vocabularies:

A controlled vocabulary is a prescribed set of consistently used and carefully defined terms, which promotes consistency in the indexing of documents and facilitation of searching (<http://www.geospatial.govt.nz/glossary/>
<http://dublincore.org/documents/usageguide/glossary.shtml>). The resources below represent several options for controlled vocabularies, but should not be considered an SI officially approved list. Besides those listed below, additional vocabularies are also available.

- Library of Congress Authorities contains preferred geographic names, personal names, corporate names, and subject terms <http://authorities.loc.gov/>
- Getty Art and Architecture Thesaurus (AAT) is a structured vocabulary relating to fine art, architecture, decorative arts, archival materials, and material culture http://www.getty.edu/research/conducting_research/vocabularies/aat/
- Library of Congress Thesaurus for Graphic Materials is a thesaurus of subject terms pertaining to topics in pictures as well as genre and format terms for photographs, prints, drawings and other graphic materials <http://www.loc.gov/rr/print/tqm2/>
- The Integrated Taxonomic Information System (ITIS) is the authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world, <http://www.itis.gov/>
- The U.S. Board on Geographic Names, <http://geonames.usgs.gov/>, is the official repository of domestic geographic names data and sanctions the standard spellings of all foreign place names. It is the official vehicle for geographic names use by all departments of the Federal Government, and the source for applying geographic names to Federal electronic and printed products. For domestic names, see the Geographic Names Information System (GNIS), <http://geonames.usgs.gov/pls/gnispublic/f?p=127:1:3019481933857537>. For foreign names, see the Geographic Names Server (GNS), <http://earth-info.nga.mil/gns/html/index.html>

Recommendation to SI Staff regarding embedded metadata when using vendors for digitization:

Material at the Institution is often outsourced to vendors for digital conversion. It is the recommendation of this group that vendors who are responsible for delivering digital surrogates be made aware of the minimal requirements for embedded metadata. It is advised that these requirements be provided in the RFP to ensure that the vendor's quote considers these requirements. It is also suggested that before the project begins,

a review of required IPTC fields and data values be discussed and agreed upon with the vendor. Any automatic descriptive or administrative data that will be delivered as part of the project from the vendor should be noted at this time. Units should provide the vendor with the exact wording for any fields the unit is requesting them to populate with data. Also insure that the vendor is not misusing IPTC /Exif fields such as embedding their company's URLs, copyright statements, scanning technician proper names, etc.

Be mindful of using templates if data is to change over the course of production. A quality assurance routine should be determined before the project begins and quality control conformance should be evaluated at regular intervals throughout the project's lifespan. This should be documented by the vendor and submitted with the digital surrogates. An acceptance/fail rate should be agreed upon before the project begins, in case it becomes necessary to resubmit files to the vendor for correction.

Data Value Types Not Recommended:

Embedded metadata by its nature is part of the image file and therefore travels with the file throughout the entire lifecycle from creation to archiving or deletion. As a result, data embedded should be information that will remain valid as long as the file exists which, for a museum, is theoretically forever. Data that change over time are better served in a rights management system, a collections information system (CIS) or other management database linked to the image. Information such as website addresses, phone numbers, emails, etc. will not remain valid in the long term and should be avoided in embedded metadata schemas.

It is also important to note that the image file will likely be repurposed or reused frequently throughout its lifetime. All embedded metadata should be acceptable for viewing by all persons, including the general public. Sensitive information, be it insurance valuations, localities of endangered species, specific storage locations, etc. would not be appropriate embedded data types.

IV. Conclusion:

As digitization becomes recognized as a critical part of the core business functions at the Smithsonian in order to support the goals and objectives of the Smithsonian Strategic Plan, we, as creators and stewards of the Institution's digital wealth, must implement procedures and processes that will make Goal Broadening Access⁶ a viable goal.

We have made great stride in the development and creation of these guidelines. But we have more to do. Other digital formats await our attention, such as audio and video files. By improving our workflow through improved standards and adoption of best practices, our digitization infrastructure will be better equipped in meeting future

⁶ SI Strategic Plan, FY 2010-2015, page 4.

measures of success, including doubling the current rate of digitization, and the quality and reuse of digital assets.⁷

As the Smithsonian begins to digitize and make its digital assets accessible, it begins to build a digital foundation for its collections and research. Accessing this digital foundation will be crucial to the Institution. As system changes are implemented, the ability to search and readily locate these digital assets becomes a key component to the SI Digitization Strategic Plan. As we begin to move forward in creating these digital assets and pushing them out to the public, it is at the creation stage that we must also have the systems search and retrieval functions in mind. These images will only be as accessible as we make them. In light of the goals and objectives of the SI Strategic Plan, a commonly understood and implemented metadata terminology would be beneficial to the Institution. It would minimize confusion in retrieving assets, readily identify which unit and collection within that unit the image came from, and also inform the user that the image may have a copyright restriction and that contacting the particular holding unit of the digital asset may be appropriate before use. This guideline has been created with the intention of educating and providing guidance for those working with digital image collections. As we established these guidelines, we have kept in mind the various needs and types of data across the institution. Our core model suggests elements that are consist across SI, while our suggested set of elements was established to accommodate the needs that one unit may have over another. The guidelines are for still image digital files. It also includes born digital material. The group has also realized the need to address PDF, video, and audio and would hope to work on each over the course of the year. The group recommends that this document be reviewed every two years to stay current with best practices.

The following people have participated in this sub group and have helped in the realization of this document:

Stephanie Ogeneski Christensen-NAA (co-leader), Doug Dunlop-SIL (co-leader), Lowell Ashley-SIL, Ricc Ferrante-SIA, Cindy Frankenburg-NMAI, Ducky Nguyen-NMAI, Kay Peterson-NMAH, Suzanne Pilsk-SIL, Ken Rahaim-OCIO, Marguerite Roby-SIA, Erin Rushing-SIL, Stephanie Smith-CFCH, Rebecca Snyder-NMNH, Amy Staples-NMAfa, Sarah Stauderman-SIA, Patti Williams-NASM

Thanks to Merry Foresta for comments and editing.

⁷ SI Strategic Plan, FY 2010-2015, page 24.

V. Appendices

Appendix A:

Element Narrative

The element narratives are meant to give a more detailed description based on the IPTC Photo Metadata. The table provided in this document gives Element Names based on these fields. This section is to help assist in understanding and ease of use of the core and suggested fields. The Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images core and suggested fields reflect how the Embedded Metadata Group has determined how the definitions below would best suit a majority of needs across the Institution. The following descriptions of the above Element Names have been taken from the IPTC Photo Metadata Specification (July 2009). Please refer to http://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata%28200907%29_1.pdf for more detailed information.

Document Title:

Definition	A shorthand reference for the item. Title provides a short human readable name which can be a text and/or numeric reference. It is not the same as Headline.
Photo Definition	A shorthand reference for the digital image. Title provides a short human readable name which can be a text and/or numeric reference. It is not the same as Headline.
Photo Help	Enter a short verbal and human readable name for the image, this may be the file name
Text	
User Note(s)	Many use the Title field to store the filename of the image, though the field may be used in many ways. Formal identifiers are provided by the Digital Image Id, or the Registry Entry property of the IPTC Extension.

Copyright Notice:

Definition	Contains any necessary copyright notice for claiming the intellectual property for this item
-------------------	--

Source:

Definition	Identifies the original owner of the copyright for the intellectual content of the item. This could be an agency, a member of an agency or an individual. Source could be different from Creator and from the entities in the Copyright Notice.
Photo Definition	Identifies the original owner of the copyright for the intellectual content of the image. This could be an agency, a member of an agency or an individual. Source could be different from Creator and from the entities in the Copyright Notice.
Photo Help	Enter the original owner of the copyright of this image
Text	

Creator:

Definition	Contains the name of the person who created the content of this item, a photographer for photos, a graphic artist for graphics, or a writer for textual news, but in cases where the photographer should not be identified the name of a company or organisation may be appropriate.
Photo Definition	Contains the name of the photographer, but in cases where the photographer should not be identified the name of a company or organisation may be appropriate.
Photo Help Text	Enter the name of the person that created this image
User Note(s)	The creator can be expressed in a more controlled way using the "Image Creator" of PLUS in the IPTC Extension additionally. It is the user's responsibility to keep the values of the IPTC Core and the PLUS fields in sync.
Implementation Note(s)	Synchronising this Creator property with PLUS' Image Creator Name should be supported by software.

Date:

Definition	Designates the date and optionally the time the intellectual content was created rather than the date of the creation of the physical representation.
Photo Definition	Designates the date and optionally the time the content of the image was created rather than the date of the creation of the digital representation.
Photo Help Text	Enter the Date the image was taken
User Note(s)	
Implementation Note(s)	If a software system requires explicit time values and no time is given by the Date Created property the software system should default the time to 00:00:00. If the software system does not require an explicit time value the time part should be left empty as it is.

Description:

Definition	A textual description, including captions, of the item's content, particularly used where the object is not text.
Photo	A textual description, including captions, of the image.
Definition	
Photo Help Text	Enter a "caption" describing the who, what, and why of what is happening in this image, this might include names of people, and/or their role in the action that is taking place

Keywords:

Definition	Keywords to express the subject of the content. Keywords may be free text and don't have to be taken from a controlled vocabulary. Codes from the controlled vocabulary IPTC Subject NewsCodes must go to the "Subject Code" field.
Photo	Keywords to express the subject of the image. Keywords may be free text and don't have to be taken from a controlled vocabulary. Codes from the controlled vocabulary IPTC Subject NewsCodes must go to the "Subject Code" field.
Definition	
Photo Help Text	Enter any number of keywords, terms or phrases used to express the subject matter in the image.
Text	

Credit/Provider:

Definition	The credit to person(s) and/or organisation(s) required by the supplier of the item to be used when published. This is a free-text field.
Photo	The credit to person(s) and/or organisation(s) required by the supplier of the image to be used when published. This is a free-

Definition text field.
Photo Help Enter who should be Credited when this image is published

Text

User Note(s) Note 1: For more formal identifications of the creator or the owner of the copyrights of this image other rights properties may be used.
Note 2: This property was named "Credit" by the IIM metadata, then it was renamed to "Provider" in IPTC Core 1.0 and has been renamed to "Credit Line" in IPTC Core 1.1. as this was actually the way the field has been used

Job Identifier:

Definition Number or identifier for the purpose of improved workflow handling. This is a user created identifier related to the job for which the item is supplied.

Photo Number or identifier for the purpose of improved workflow handling. This is a user created identifier related to the job for which the image is supplied.

Definition

Photo Help Enter a number or identifier needed for workflow control or tracking
Text

Headline:

Definition A brief synopsis of the caption. Headline is not the same as Title.

Photo Definition A brief synopsis of the caption. Headline is not the same as Title.

Photo Help Text Enter a brief publishable synopsis or summary of the contents of the image

Appendix B: Examples of IPTC Fields Mapped to Various Imaging Programs

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

Appendix: IPTC Fields Mapped to Various Imaging Programs

Yellow= Core Green= Suggested Gray= Not Suggested

IPTC		Adobe Photoshop v6		Adobe Photoshop v7		Adobe P-Shop CS (v8)		Adobe P-ShopCS2 (v9)		Adobe P-Shop CS3 (v10)	
IPTC IIM (v4)	size	IPTC Core (v1)	size	Object Name	Title	Document Title	Document Title	Document Title	Document Title	Title	
Object Name (Title)	64	Title	64	Object Name	Title	Document Title	Document Title	Document Title	Document Title	Title	
Urgency	1	deprecated		Urgency	Urgency	Urgency	Urgency	Urgency	Urgency		
Category	3	deprecated		Categories	Categories	Categories	Categories	Categories	Categories		
Supplemental Categories	32	deprecated		Supplemental Categories	Supplemental Categories	Supplemental Categories	Supplemental Categories	Supplemental Categories	Supplemental Categories		
Keywords	64	Keywords	64	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	
Special Instructions	256	Instructions	256	Special Instructions	Instructions	Instructions	Instructions	Instructions	Instructions	Instructions	
Date Created	8	Date Created	8	Date Created	Date Created	Date Created	Date Created	Date Created	Date Created	Date Created	
By-line (Author)	32	Creator	32	By-line	Author	Author	Author	Author	Author	Creator	
By-line Title (Author Position)		Creator's Title	32	By-line Title	Author's Position		Author Title	Author Title	Author Title	Creator Job Title	
City	32	City	32	City	City	City	City	City	City	City	
Province/State	32	Province/State	32	State/Province	State/Province	State/Province	State/Province	State/Province	State/Province	State/Province	
Country/Primary Location	64	Country	64	Country Name	Country	Country	Country	Country	Country	Country	
Original Transmission Reference	32	Job Identifier	32	Original Transmission Reference	Transmission Reference	Transmission Reference	Transmission Reference	Transmission Reference	Transmission Reference	Job Identifier	
Headline	256	Headline	256	Headline	Headline	Headline	Headline	Headline	Headline	Headline	
Credit	32	Provider	32	Credit	Credit	Credit	Credit	Credit	Credit	Provider	
Source	32	Source	32	Source	Source	Source	Source	Source	Source	Source	
Copyright Notice	128	Copyright Notice	128	Copyright Notice	Copyright Notice	Copyright Notice	Copyright Notice	Copyright Notice	Copyright Notice	Copyright Notice	
Caption/Abstract	2000	Caption/Description	2000	Caption	Caption	Description	Description	Description	Description	Description	
Writer/Editor	32	Caption/Description Writer	32	Caption Writer	Caption Writer	Description Writer	Description Writer	Description Writer	Description Writer	Description Writer	
		Creator's Contact Info (CCI): City						Creator City		Creator City	
		CCI: Country						Creator Country		Creator Country	
		CCI: Address						Creator Address		Creator Address	
		CCI: Postal Code						Creator Postal Code		Creator Postal Code	
		CCI: State/Province						Creator State/Province		Creator State/Province	
		CCI: Email(s)						Creator Email(s)		Creator Email(s)	
		CCI: Phone(s)						Creator Phone(s)		Creator Phone(s)	
		CCI: WebURL(s)						Creator WebURL(s)		Creator WebURL(s)	
		ISO Country Code	3					Country Code		Country Code	
		Intellectual Genre						Intellectual Genre		Intellectual Genre	
		Location	32					Location		Location	
		Rights Usage Terms						Rights Usage Terms		Rights Usage Terms	
		Subject Code						IPTC Subject Code		IPTC Subject Code	
		IPTC Scene						IPTC Scene		IPTC Scene	
Edit Status	32										
Fixture Identifier	32				Job Name						
Content Location Name	64										
Contact	128										
				Image URL	Owner URL	Copyright Info URL	Copyright Info URL	Copyright Info URL	Copyright Info URL	Copyright Info URL	
				Mark as Copyrighted	Copyright Status	Copyright Status	Copyright Status	Copyright Status	Copyright Status	Copyright Status	
Fields in blue are part of the IPTC IIMv4 schema, and used by iView, but not visible by any other programs except via plug-ins											
Fields in red will not be read in any programs except Photoshop, and/or other programs using the Adobe "File Info" data format											

** This table is courtesy Ken Rahiam

Adobe	Adobe	Adobe	Adobe	Apple	Camera Bits	iView	iView
P-ShopCS4 (v11)	P-Shop CS3 (v10)	P-ShopCS4 (v11)	P-Shop Elements3	Aperture v1.5	PhotoMechanic 4.5	MediaPro v2.6	MediaPro v3.1
Title	Title	Title	Document Title	Object Name	Object Name	Product	Title
				Uragency	Uragency	Color Label	Uragency
			Categories	Category	Category	Genre	Category
			Supplemental Categories	Supplemental Categories	Supp Cat1/Cat2/Cat3	Categories	Supplemental Categories
Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords
Instructions	Instructions	Instructions		Special Instructions	Special Instructions	Instructions	Instructions
Date Created	Date Created	Date Created		Date Created	Date Created	Event Date	Date Created
Creator	Creator	Creator	Author	Byline	Photographer	Author	Creator
Creator Job Title	Creator Job Title	Creator Job Title		Byline Title	Title	Author Title	Creator's Job Title
City	City	City		City	City	City	City
State/Province	State/Province	State/Province		Province/State	State	State	State/Province
Country	Country	Country		Country Name	Country	Country	Country
Job Identifier	Job Identifier	Job Identifier		Original Transmission	Transmission Ref	Transmission	Job Identifier
Headline	Headline	Headline		Headline	Headline	Title	Headline
Provider	Provider	Provider		Credit	Credit	Credit	Provider
Source	Source	Source		Source	Source	Source	Source
Copyright Notice	Copyright Notice	Copyright Notice	Copyright Notice	Copyright Notice	Copyright	Copyright	Copyright Notice
Description	Description	Description	Description	Description	Caption	Caption	Description
Description Writer	Description Writer	Description Writer	Description Writer	Writer/Editor	Caption Writers	Writer	Description Writer
Creator City	Creator City	Creator City			Contact City		City
Creator Country	Creator Country	Creator Country			Contact Country		Country
Creator Address	Creator Address	Creator Address			Contact Address		Address
Creator Postal Code	Creator Postal Code	Creator Postal Code			Contact Zip		Postal Code
Creator State/Province	Creator State/Province	Creator State/Province			Contact State		State/Province
Creator Email(s)	Creator Email(s)	Creator Email(s)			Contact Email(s)		Email(s)
Creator Phone(s)	Creator Phone(s)	Creator Phone(s)			Contact Phone(s)		Phone(s)
Creator WebURL(s)	Creator WebURL(s)	Creator WebURL(s)			Contact Web URL(s)		Website(s)
Country Code	Country Code	Country Code			Code		Country Code
Intellectual Genre	Intellectual Genre	Intellectual Genre			Intellectual Genre		Intellectual Genre
Location	Location	Location			Location		Location
Rights Usage Terms	Rights Usage Terms	Rights Usage Terms			Rights Usage Terms		Rights Usage Terms
IPTC Subject Code	IPTC Subject Code	IPTC Subject Code			IPTC Subject Code		Subject Code
IPTC Scene	IPTC Scene	IPTC Scene			IPTC Scene		Scenes
				Edit Status			Edit Status
				Fixture Identifier			Event
				Contact Location Name			(IIM) Location
							People
Copyright Info URL	Copyright Info URL	Copyright Info URL	Copyright Info URL		Copyright Info URL		
Copyright Status	Copyright Status	Copyright Status	Copyright Status		Copyright Status		
				Version Name			
				Contact Location Code			
				Editorial Update			
				Expiration Date			
				Image Orientation			

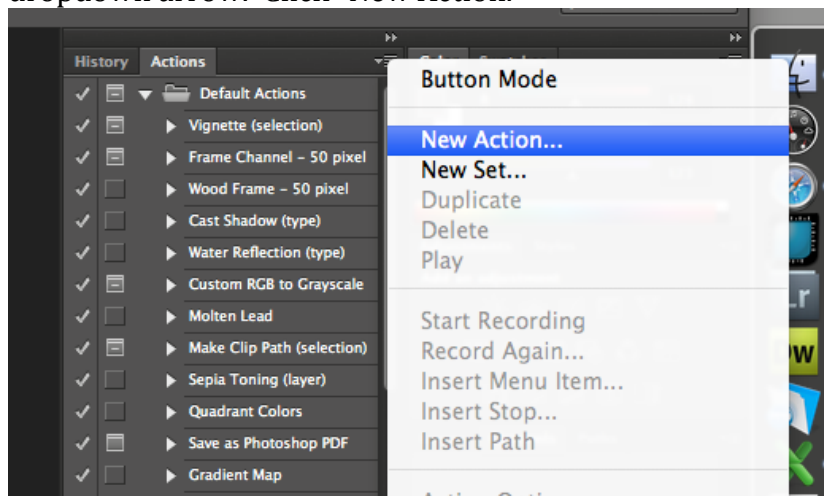
Quick checklist for processing images—NAA Digital Imaging Lab

Once images have been digitized, there are steps for quality control and post processing. Many of the following steps can be automated with batch actions in Photoshop. Below is a checklist of steps that need to be completed before derivative files (PDFs and/or jpgs) can be created and before images can be ingested into the DAMS:

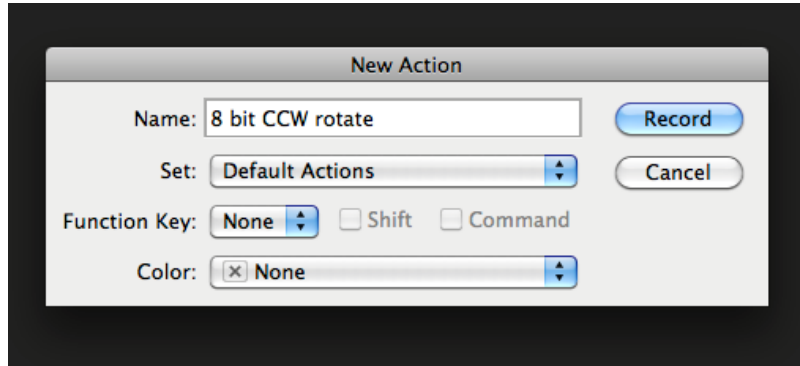
1. **Convert 16-bit tiffs to 8-bit mode**
2. **Rotate images (if necessary)**
3. **Embed descriptive metadata (if necessary)**
4. **De-skew images as needed**
5. **Crop images**
6. **Inspection of 10% of images at 100% magnification**

1. **Convert 16-bit tiffs to 8-bit mode**

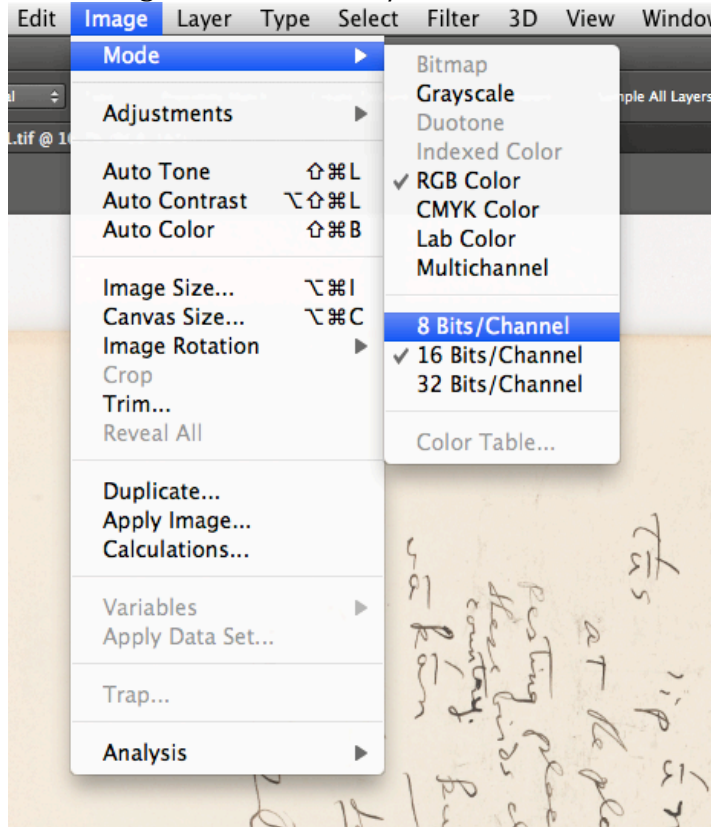
- a. Digitized images are captured in a raw .mos file and then processed to 16-bit tiffs in the Leaf Capture software. Those 16-bit files need to be converted to 8-bit to make them easier to work with. This can be automated with a batch action in Photoshop.
- b. Open Photoshop. Go to the Action tab on the right side and click the dropdown arrow. Click “New Action.”



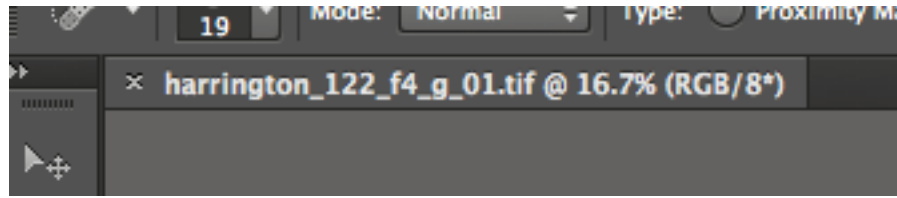
A new naming window will pop up. Create a simple, explanatory name for the action, such as “8 bit and CCW rotate.”



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

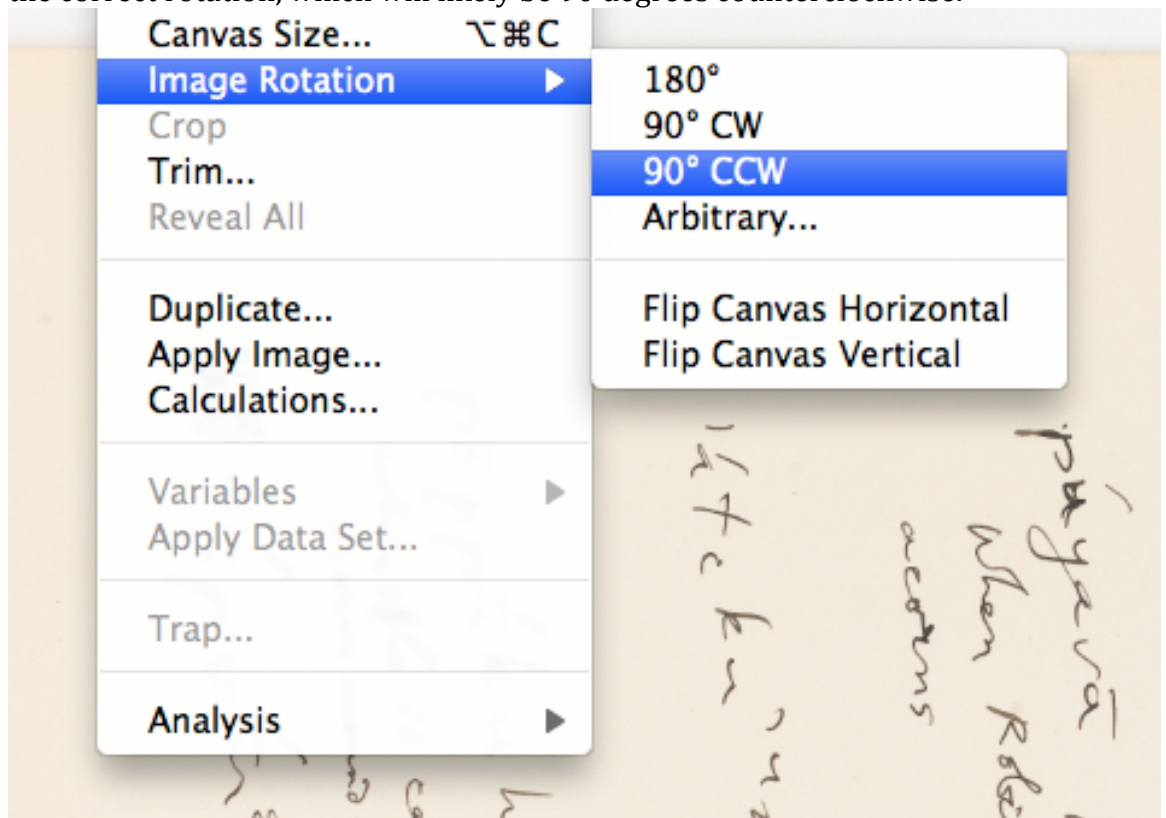


- f. Once you have converted the image to 8 Bits/Channel, you may notice that Photoshop will reflect that in the tab where the filename of the image is (RGB/8* indicates this):



2. Rotate images (if necessary)

- a. If images are not shot "right side up" they will need to be rotated. This can also be created in a Batch Action. Follow the same steps for creating a batch action above, but instead of converting the mode, you will be rotating the images. And in fact, converting to 8-bit and rotating images can be done in the same batch action. After you have completed the above steps, go to Image→Image Rotation and choose the correct rotation, which will likely be 90 degrees counterclockwise.

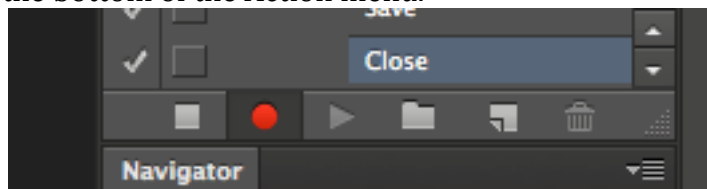
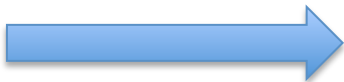


After you have done this, you will notice the image is now "right side up."

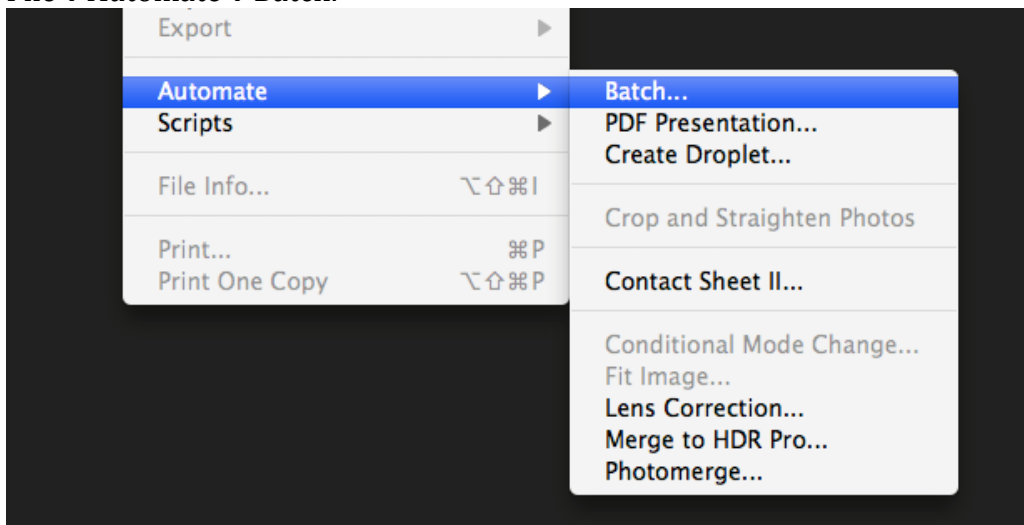
- b. Go to File→Save. Once the dropdown menu has disappeared, go to File→Close. If you get the following popup window, just click Ok.



- c. 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.

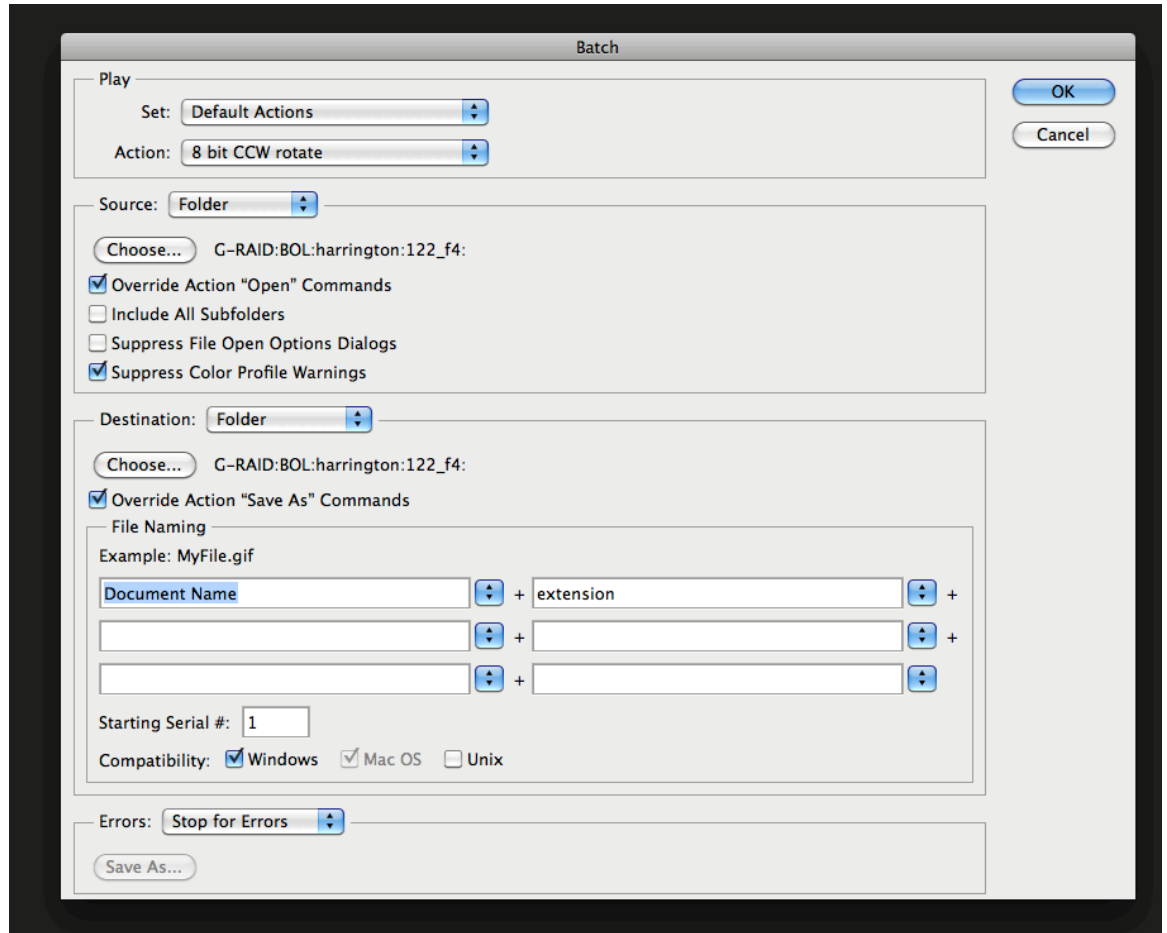


- d. Now that you have created the Action, you can “play” it so the steps are applied to the entire folder of images. Go to File→Automate→Batch.



- e. 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:

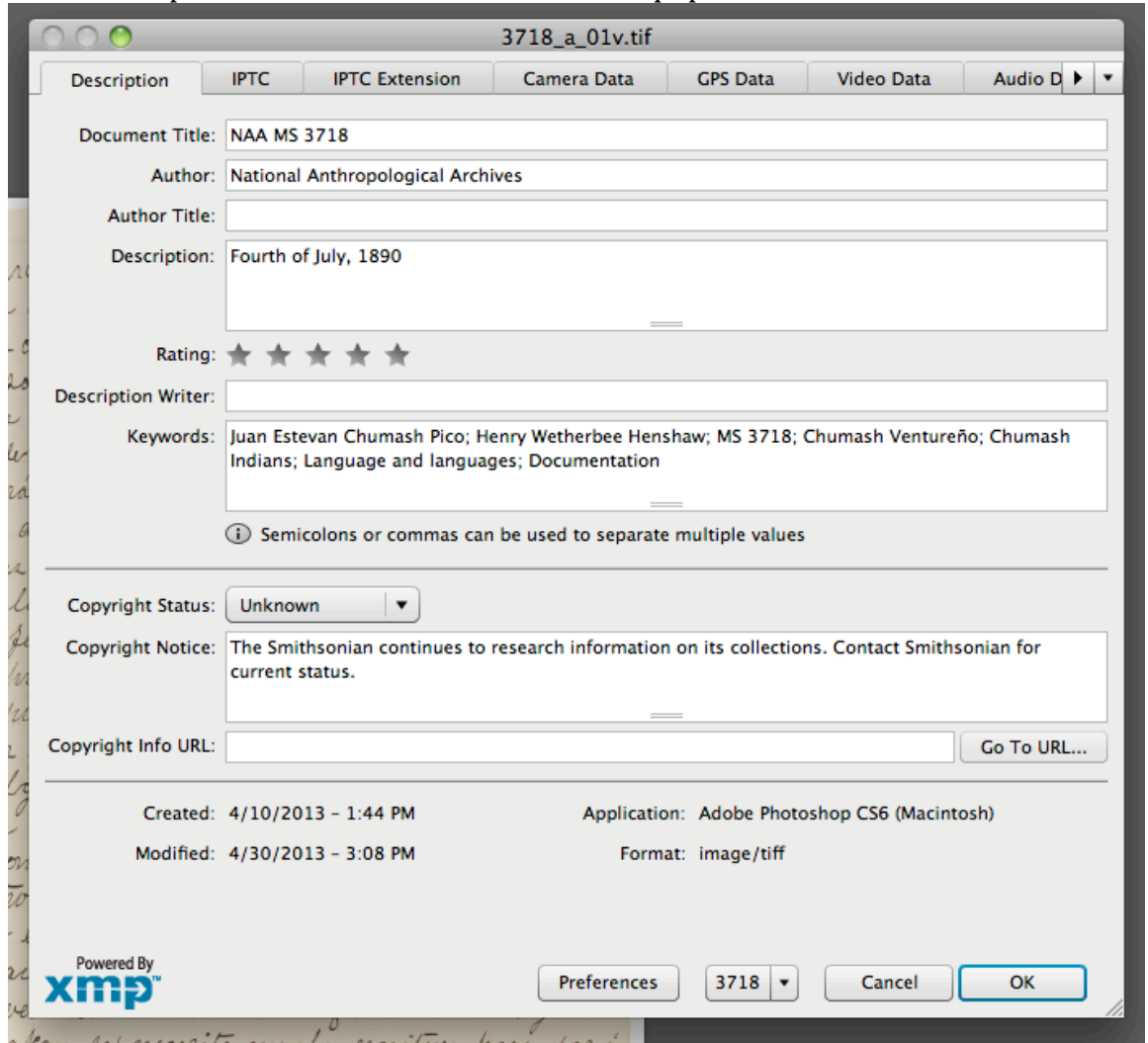


- f. Click “Ok.” The batching will automatically begin and run through the files contained within the folder, which will take anywhere from a few seconds to a few minutes, depending on the amount and size of the files.

3. Embed descriptive metadata (if necessary)

- a. The newer version of the Leaf Capture software has the capability to embed metadata during digitization; please see “Instructions for Using the Leaf Aptus 75 Camera Back and Software” in the binder. If metadata hasn’t been added, it can be batch embedded in Photoshop or Bridge.

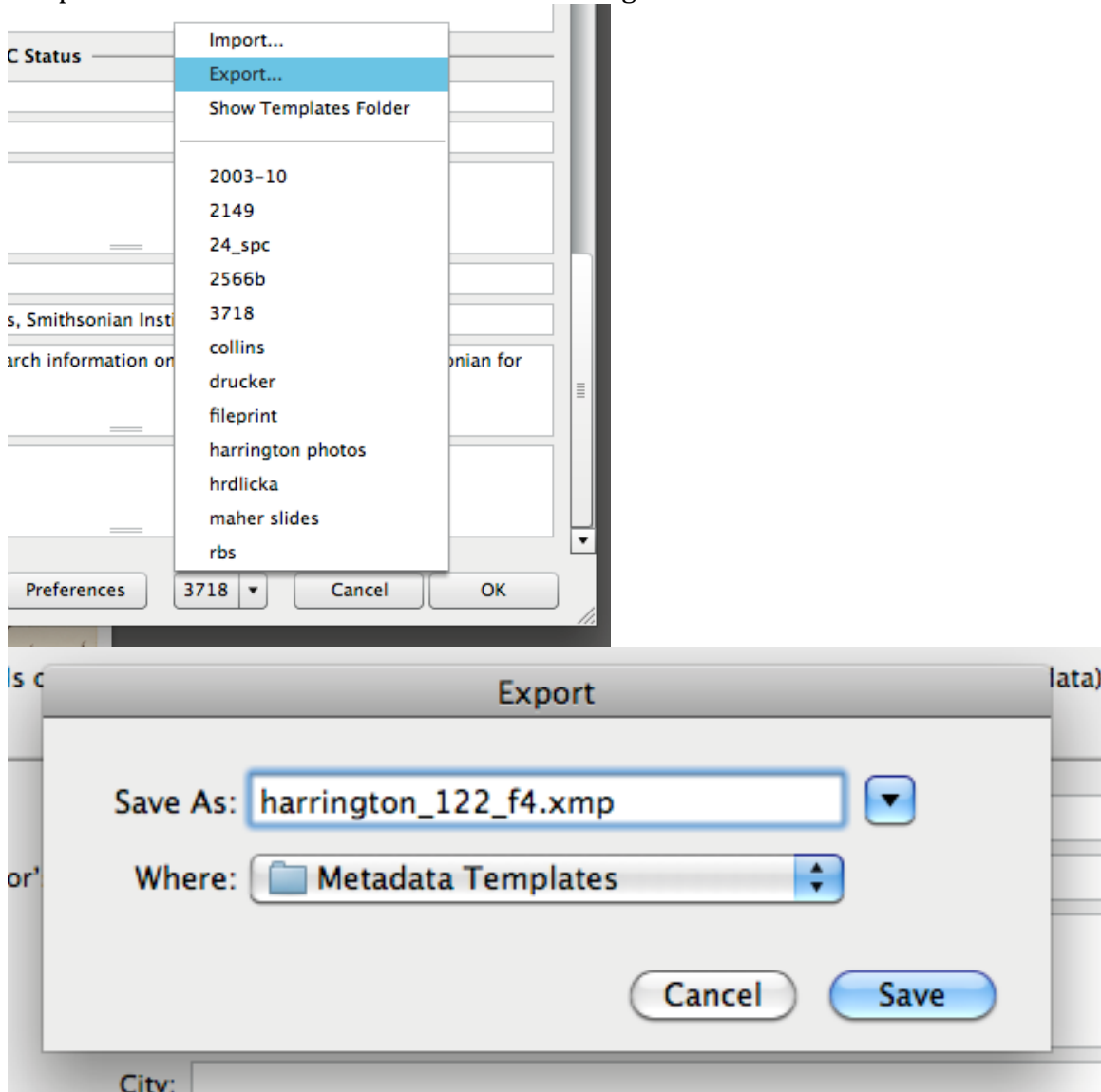
- b. Open a file from the folder you're working on, and go to File→File Info in Photoshop. This is where metadata fields are populated.



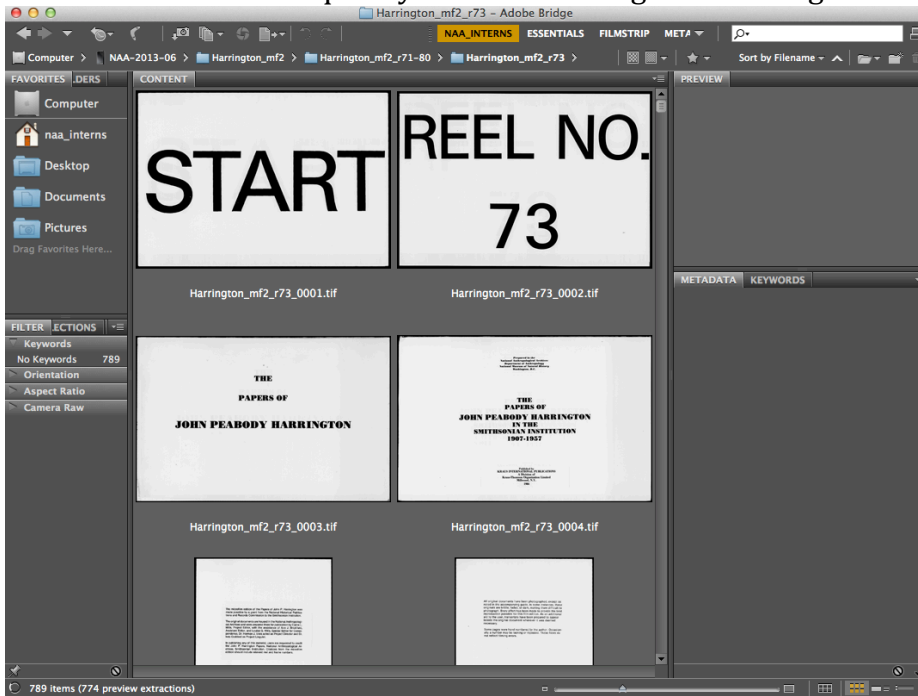
- c. Using the SIRIS online catalog, the following fields need to be populated:
- i. Document Title: The number of the manuscript (from the Local Number field in SIRIS).
 - ii. Author: National Anthropological Archives
 - iii. Description: The title of the manuscript, also from the SIRIS record.
 - iv. 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
 - v. Copyright Status: UNKNOWN. It is extremely important that the copyright status says UNKNOWN.

- vi. Copyright Notice: “The Smithsonian continues to research information on its collections. Contact Smithsonian for current status.” This is the same for all material.
- vii. In the IPTC tab, make sure the Source field is populated. It should say, “NAA-Natl. Anthropological Archives, Smithsonian Institution.”

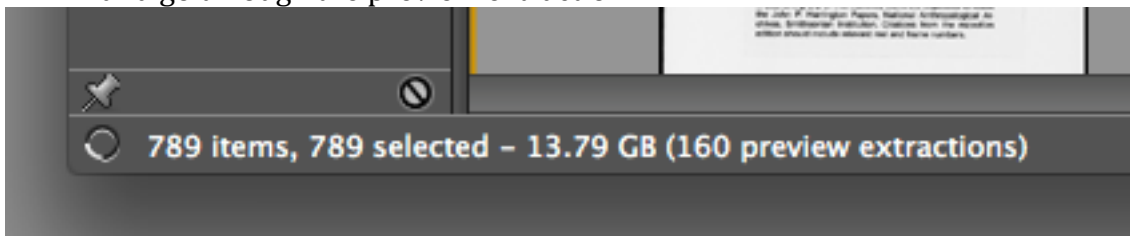
Once you have created the metadata template, you will want to Save (or Export) it for later use. In the metadata window, click Export. A new Save window will pop up, and it will automatically fill in the filename of the file currently open. You may leave it as such, or make it shorter and simpler, like 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.



Metadata can be more quickly embedded using Adobe Bridge CS6.

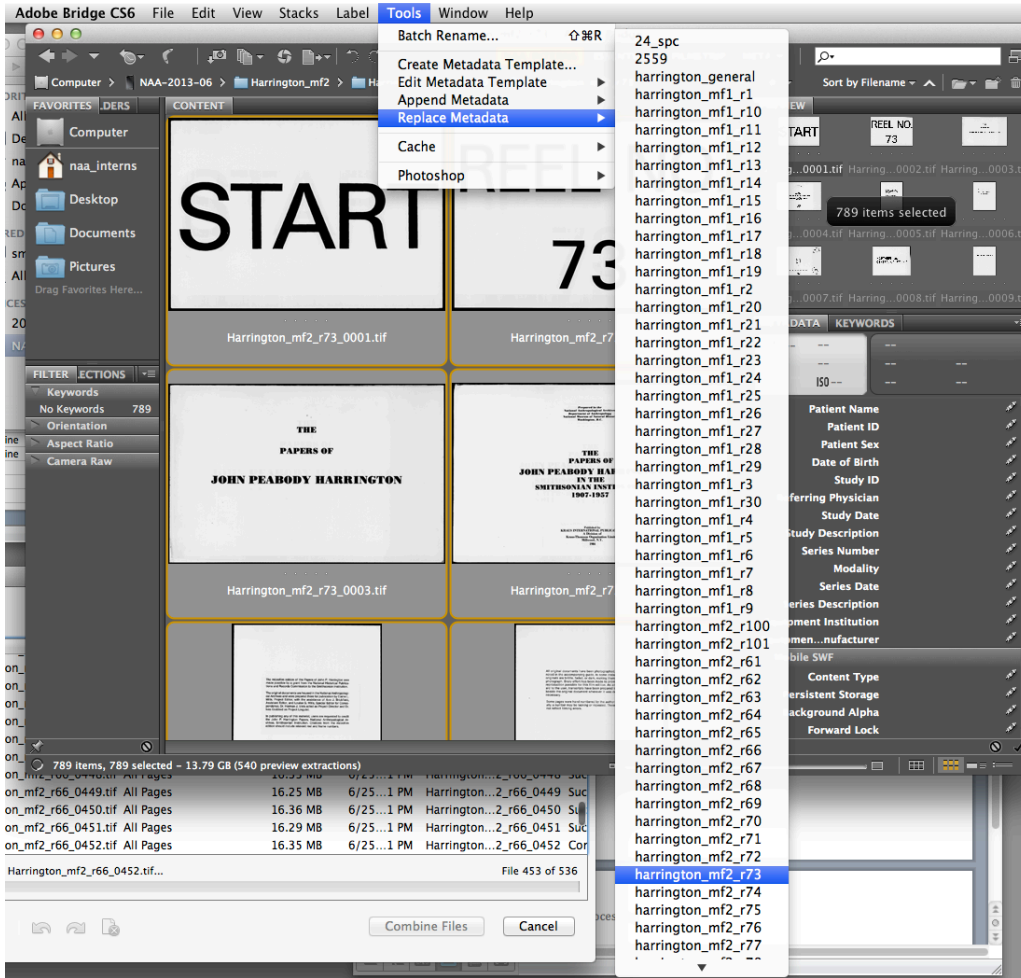


1. Open the files in Adobe Bridge. The files may take several minutes to load and go through the preview extraction.

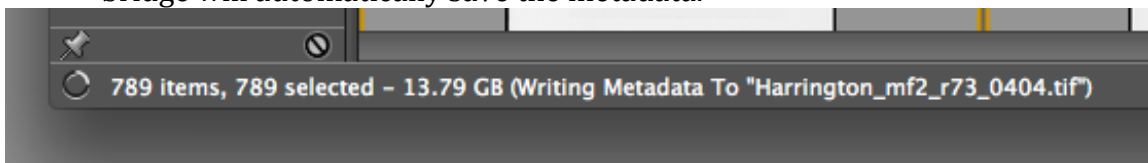


2. Once all the files have loaded, select all files (command + A) then go to “Tools” → “Replace Metadata”

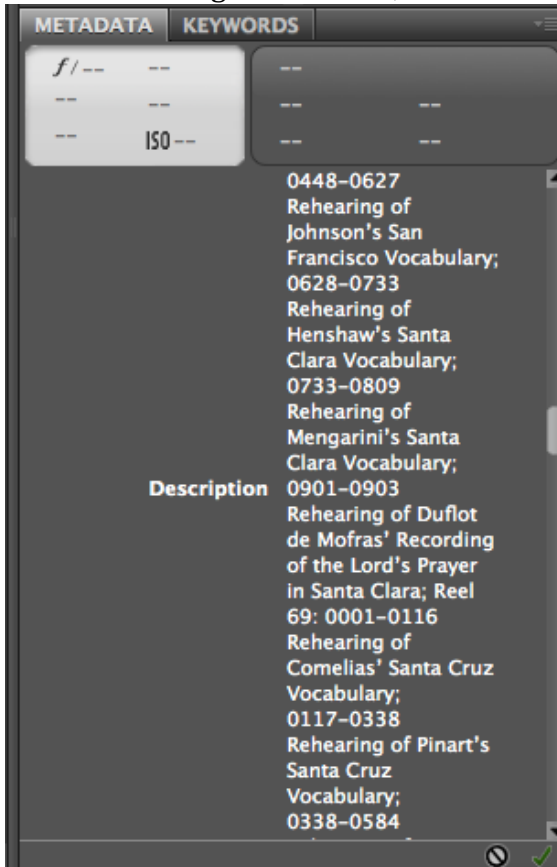
NAA Processing Guidelines September 2013



3. 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.



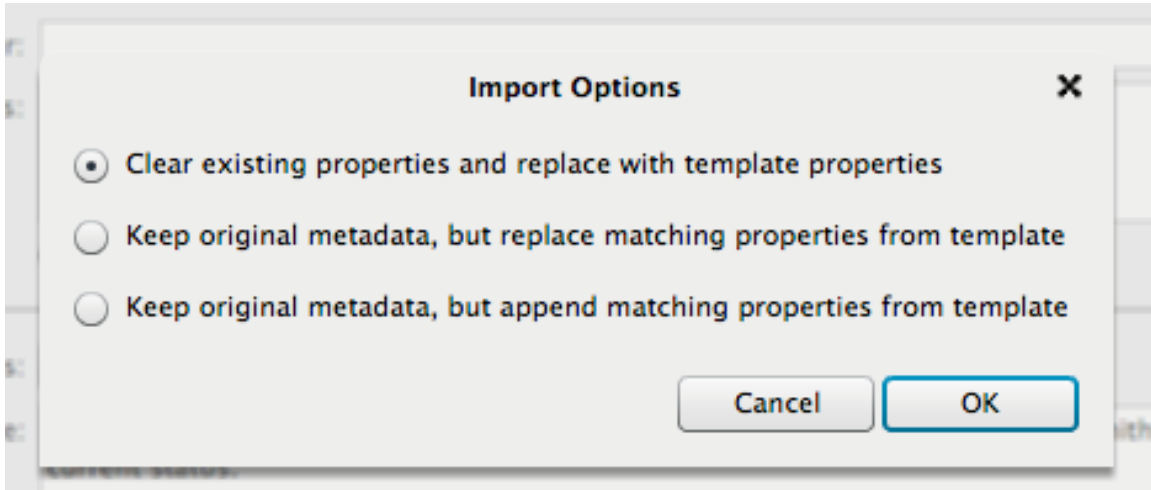
On the lower right hand side, confirm that the metadata has been embedded.



4. Confirm that metadata has been embedded by opening the files in Adobe Photoshop and checking the "File Info."

If for some reason, embedding metadata does not work in Bridge, it can be batch embedded in Photoshop. Once the metadata template has been created, you can create a Batch Action to embed the metadata to a given collection of images. Follow the steps above for creating an action. New Action → Name (should be the manuscript number + metadata so you know what information is being embedded) → Record.

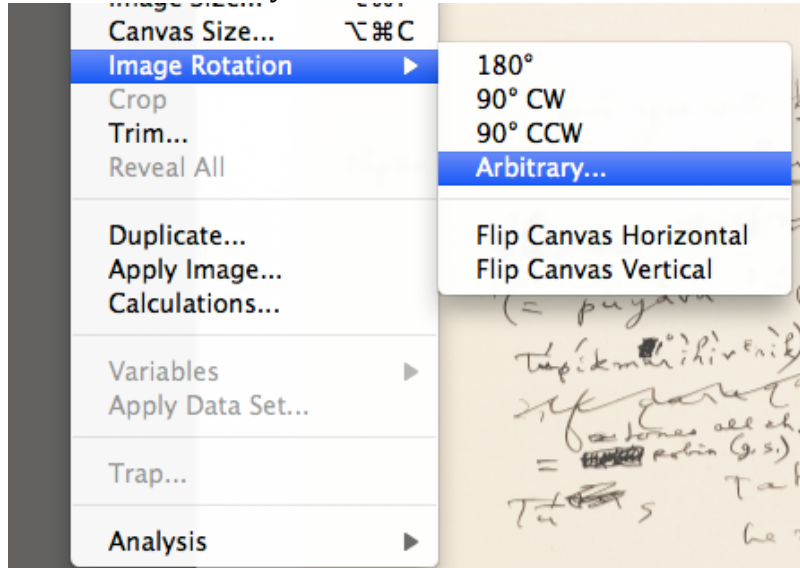
Go to File → File Info. Click the "Export" button and select the correct metadata template from the dropdown menu. When the option menu appears, click the first option:



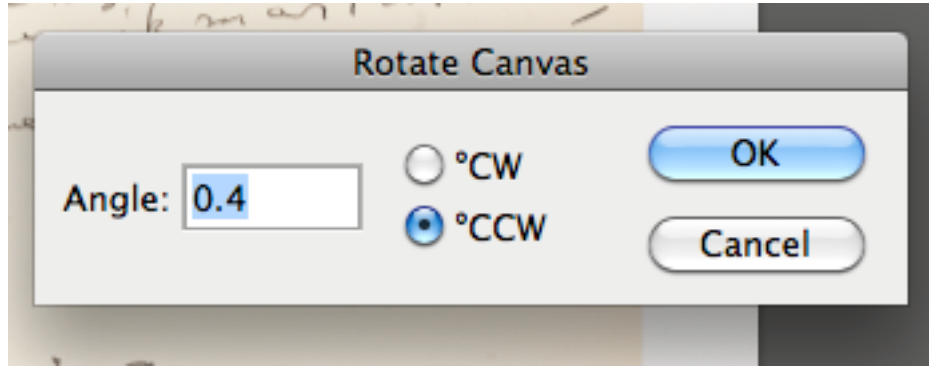
Click Ok. Then Click Ok in the metadata window. Once the window disappears, go to File→Save then File→Close. Stop action. Then go to File→Automate and follow the above steps for running action.

4. De-skew images as needed

- a. Now that all the images have to be opened individually to be de-skewed and cropped. Open the first image in the folder you are working on; if the page appears “crooked” you will need to rotate it by .2 or .4 of a degree to make it straight. Go to Image→Image Rotation→Arbitrary



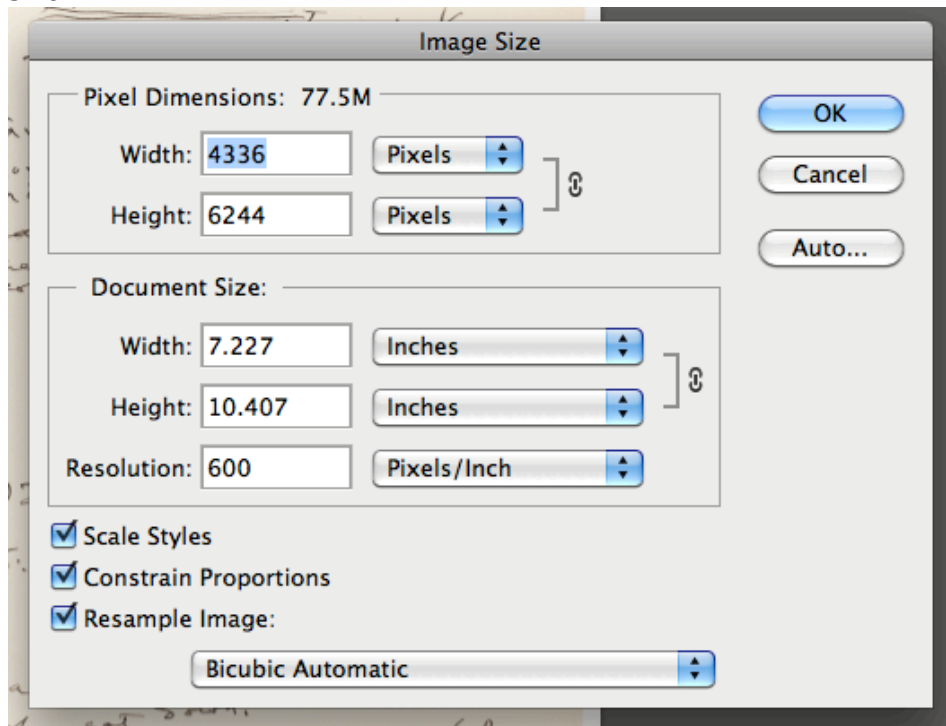
Depending on which way the images need to be de-skewed, you will choose .2 or .4 and either clockwise or counterclockwise. Click Ok.



This step may have to be repeated a few times, depending on how skewed the images is; determine this just by “eyeballing” how straight the image appears.

5. Crop images

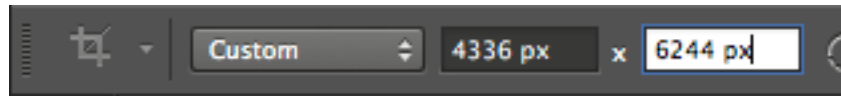
- a. The color bar needs to be cropped out of the final image for manuscript pages. Open the first image in the folder you’re working from. Select the crop tool from the tool bar on the left side. Crop the image to approximately 1/8th of an inch on all 4 sides. Make sure the white background is still visible. 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.



- b. Once the image has been cropped, you can save and close it. When you open the next images, follow the steps for de-skewing it. When you

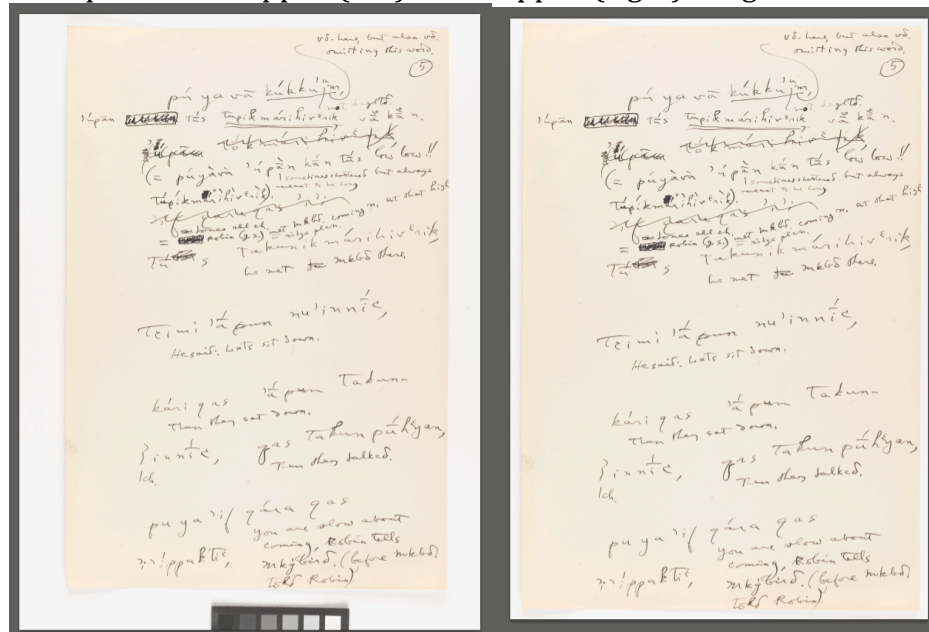
are ready to crop it, instead of cropping it “freehand” you will set the crop area based on the pixel information from the previous image.

- c. Select the crop tool. In the upper left corner of the Photoshop window, you will see the crop tool icon, a dropdown menu and a box where values can be inserted. Put your cursor into the box and enter the pixel information from the previous image size. You will put in the number followed by PX, otherwise Photoshop will tell you the size is inapplicable. There are two boxes to fill (one for width and one for height).



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.

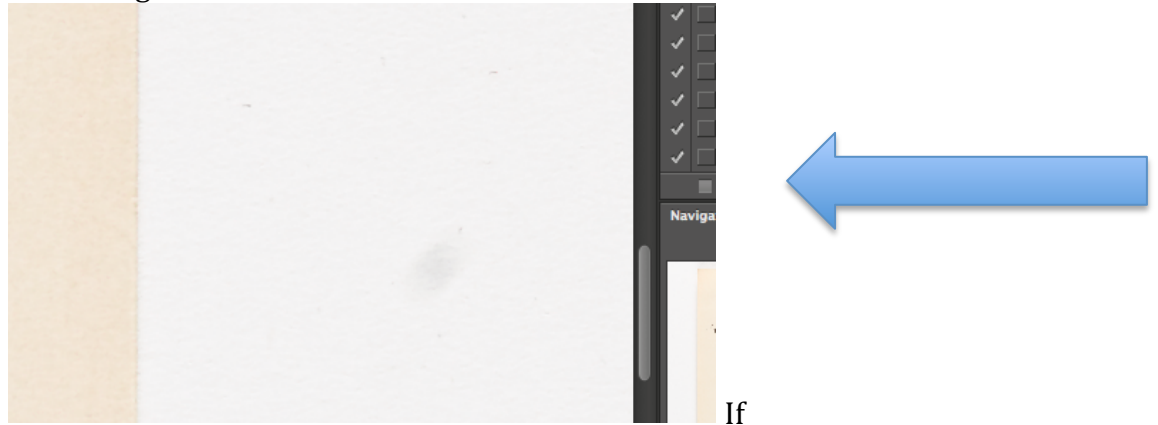
- d. Examples of uncropped (left) and cropped (right) images:



6. Inspection of 10% of images at 100% magnification

- a. At a minimum, 10 images or 10% of each batch of digital images, whichever quantity is larger, should be inspected at 100% magnification for potential defects, like dust and noise.
- b. Select the “band-aid” tool (actually called the Spot Healing Brush tool) from the toolbar on the left. At the bottom of the Photoshop window, there is a percentage at which you are viewing the image. Place your cursor in the box and change the percentage to 100%. The image will magnify. You will want to scroll your cursor over the entire image;

starting in the upper left hand corner and moving across and down is a simple way to do so. You are inspecting the image for the introduction of any dust or digital “noise” during digitization. If you spot any of these types of defects, you will use the “band-aid” to remove it. **YOU WILL NOT REMOVE ANY “HISTORICAL” IMPERFECTIONS OR DEFECTS.** The historical integrity of the image is to be maintained. Therefore, you are often looking for dust or smudges on the white background. Below is an example of a “smudge” that can be removed from the image. Use the Healing Brush to “erase” the smudge.



If you find that many of the images you are viewing at 100% appear to have defects, you may want to inspect more images (20-25%). But if many of the images appear to be clean, 10% inspection is acceptable.

Once all the images have been processed, derivative files (PDFs and JPGS) need to be created. Please see “NAA Guidelines for Embedding Metadata and Creating JPEG Derivatives,” as well as “NAA PDF Creation and Workflow” for continued instructions. If at any time you have any questions about the above workflow, **PLEASE ASK THE DIGITAL STAFF!**