6-12-2012

Optimized Still Image Batch Processing of Special Collections Bound Monographs and Manuscripts Using DNG, JPEG 2000, and Embedded XMP Metadata

Michael J. Bennett
University of Connecticut - Storrs, michael.bennett@uconn.edu

Optimized Still Image Batch Processing of Special Collections Bound Monographs and Manuscripts Using DNG, JPEG 2000, and Embedded XMP Metadata

Michael J. Bennett; University of Connecticut Libraries; Storrs, CT/USA

Abstract

Batch still image processing is examined in the context of operational bound monographs and manuscripts reformatting. The scaling of overall workflows through the flexible use of Lightroom, Photoshop, VueScan, and Jhove on parametrically-edited raw DNG and batch-rendered JPEG 2000 files is surveyed. Potential gains in processing efficiency, in comprehensive device data capture and preservation, in adaptable master image repurposing capabilities, and in the smoother growth of the required large-scale digital storage capacities that surround such operational conversions are considered.

Introduction

Digital still image capture of archives and special collections' objects has often followed a traditional uncompressed TIFF archival copy > compressed JPEG access copy processing chain for many reformatting projects. Though this has operated well enough in most cases, newer image formats and metadata wrappers along with more powerful tools centered on such advances have allowed for novel image utilization and the re-evaluation of overall workflow efficiencies. In an ever expanding electronic environment, users are in search of richer digital content and have come to expect greater image quality for innovative manipulations and enhanced study. Within this ecosystem the obligations of content creators towards coherent production, storage, management, preservation, and more flexible and finely-tailored output of their own quickly growing digital archives and special collections have become magnified as a product of increasing overall scale. In turn, it naturally follows that novel value-added enhancements in workflow design, using the inherent capabilities of new still imaging formats, metadata specifications, and the latest developments in image editing software are engineered.

DNG as RAW Safety Master File Format

When looking at raw image formats as the starting point of an overall digital imaging chain a number of scalable advantages over traditional TIFF-based archiving and raster processing become apparent. Though these are outlined in narrative depth elsewhere [1][2][3][4][5][6] a look at the current capture workflow of monographs and manuscripts employed at the University of Connecticut (UConn) Libraries may be pertinent.

Bound Monograph Workflow: DNG from Camera Color Filter Array (CFA) [7][8] Sensor Data

In this example, page images of John Donne's 1611 *Conclave Ignati* are used. Proprietary Canon .CR2 camera raw files are first converted into a folder of DNG safety masters, segregated into left and right page Adobe Lightroom 3 Collections by either verso or recto page origin, and then losslessly rotated and cropped through synchronized Lightroom parametric [9] edits. Such DNG raw editing, particularly across large, homogeneous image groups, saves substantial processing time, overall CPU overhead, and required storage space against comparable raster image batch editing steps which, unless accomplished as unmerged layered TIFF or PSD files, are irreversible in final form. Raw DNGs can be losslessly compressed, can retain originally-captured sensor data even when parametrically edited, and in fact can quite easily be reversed back to their original latent, unedited state. In this manner, the format can adroitly serve as both a robust master and efficiently processed format.

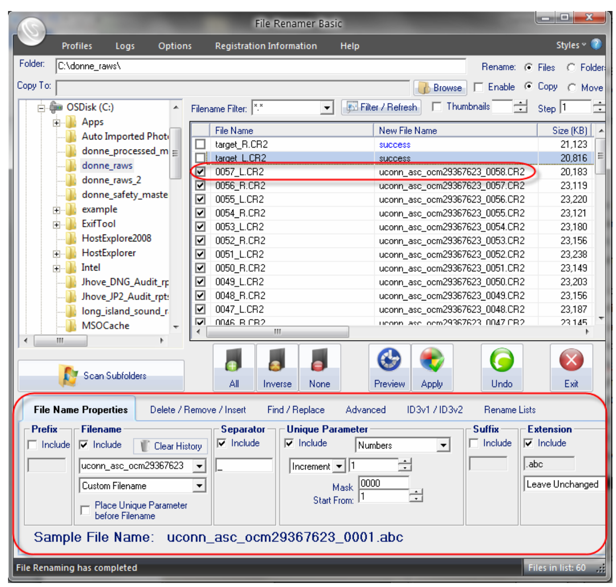


Figure 1. *Camera raw files are renamed with local file naming convention [10]. This can be done in either Adobe Bridge, or in a dedicated renaming tool like FileRenamer.*

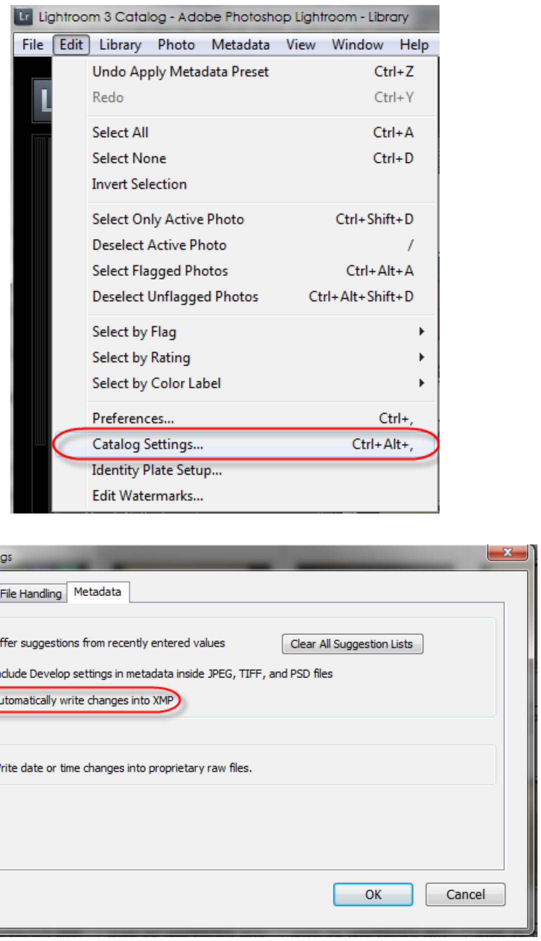


Figure 2. *Adobe Lightroom is opened and Catalog Preferences are set to "Automatically write changes into XMP." This ensures that all subsequent parametric edits and added process metadata will be embedded in the newly-converted DNG files (see Figure 3) and not just stored in the Lightroom catalog database.*

Figure 3. *A new folder, c:\book_title_safety_masters, is created. In Lightroom c:\book_title_raw (original camera raw images folder) is imported. During this process original camera raw files are batch converted to DNG and saved to a new destination folder, c:\book_title_safety_masters. By default Lightroom accomplishes such DNG conversions (v1.3) with the lossless compression supported by the format.*

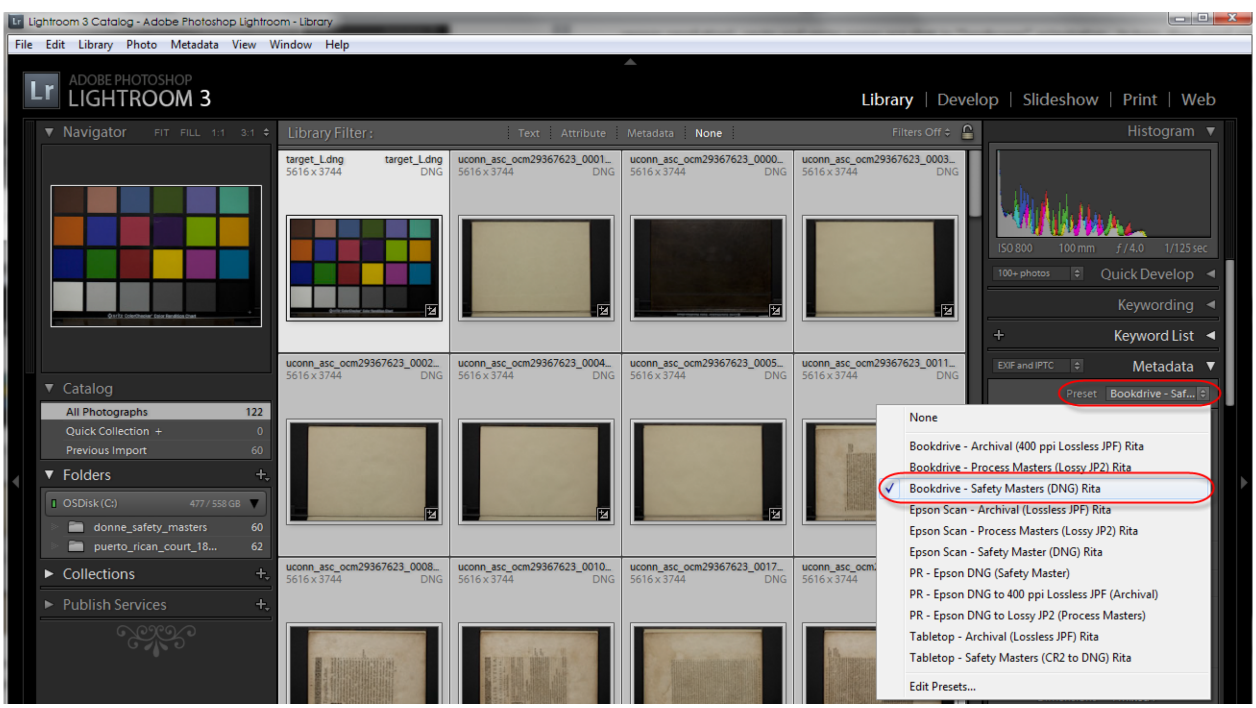
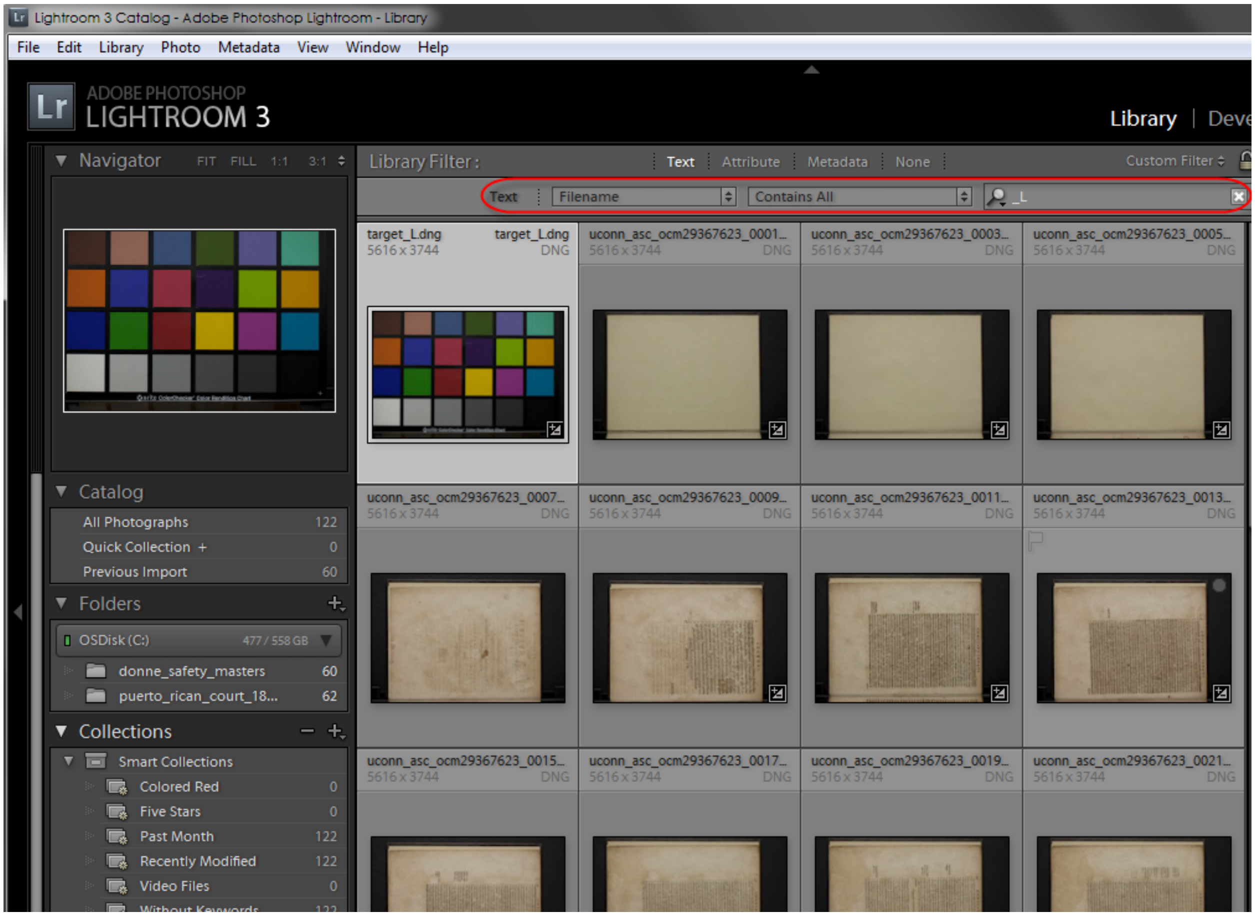
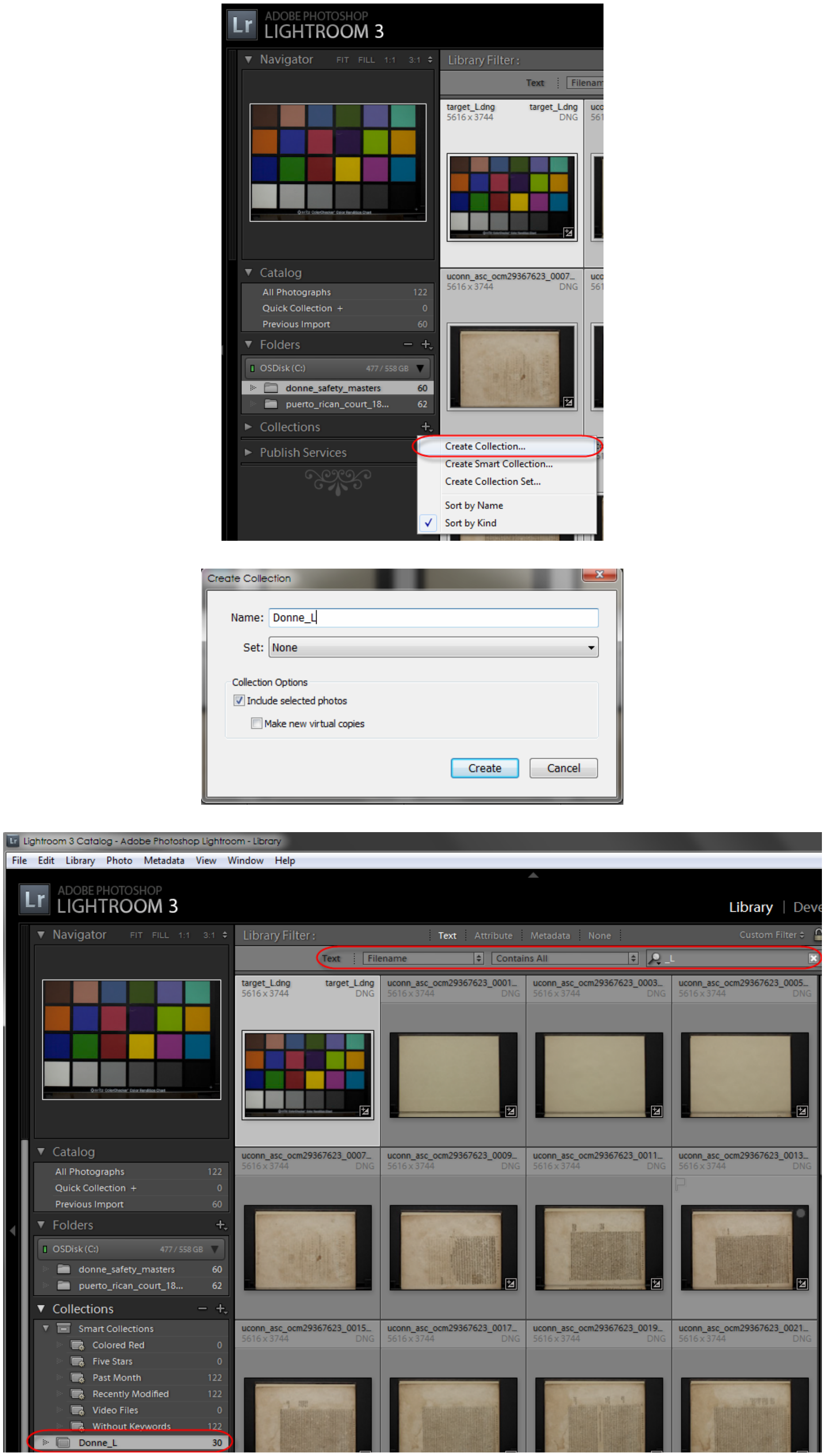


Figure 4. *With all images selected, technical process metadata is added in a batch from a pre-made Lightroom metadata preset (see section V. for more details on the various process metadata templates used in the UConn Libraries lab).*

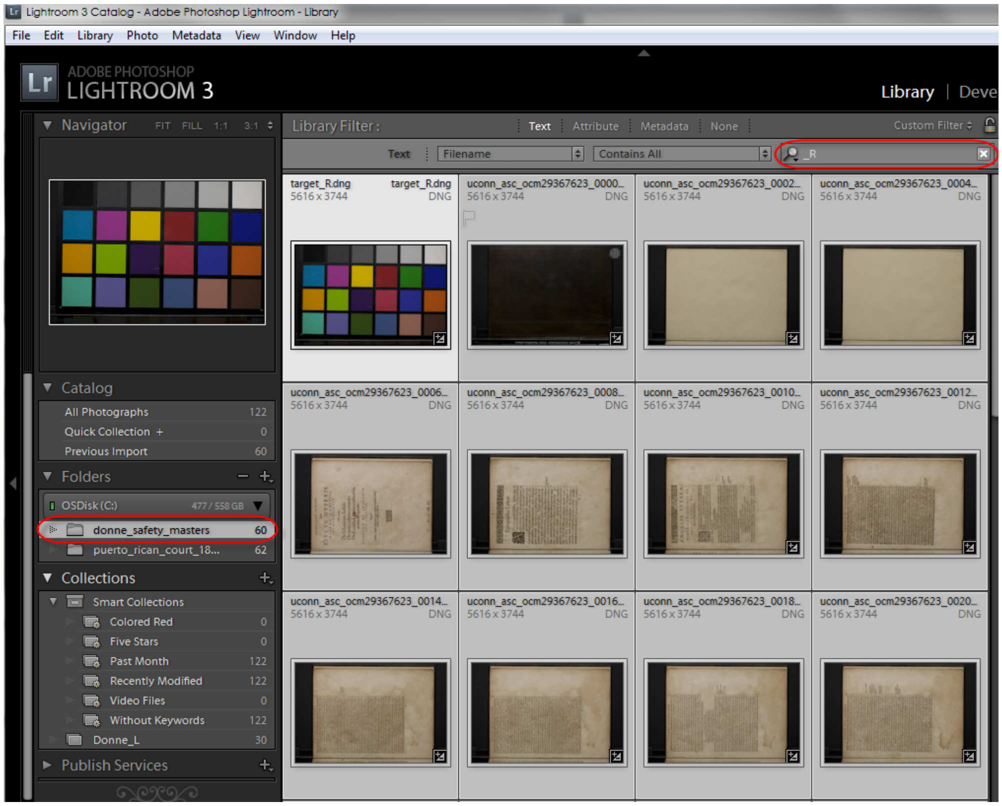
At UConn, bound monographs are captured on Atiz BookDrive book cradles outfitted with dual Canon 5D II DSLR full-frame sensor cameras that shoot 3:2 aspect ratio images. As a result in order to minimize cropping (and the loss of maximum sensor sampling rate), recto and verso pages are shot in "landscape" orientation. In turn, they require either 90° clockwise or 90° counter-clockwise rotation to bring page text back into proper "portrait" reading alignment. To best facilitate batch processing, then, left and right images are captured with _L and _R file name suffixes respectively through Atiz BookDrive Capture software. Lightroom can then easily filter by file name suffix and segregate images into left and right image collections where batch clockwise or counter-clockwise rotation and cropping steps can be parametrically run on the DNG files in a quick, lossless manner.

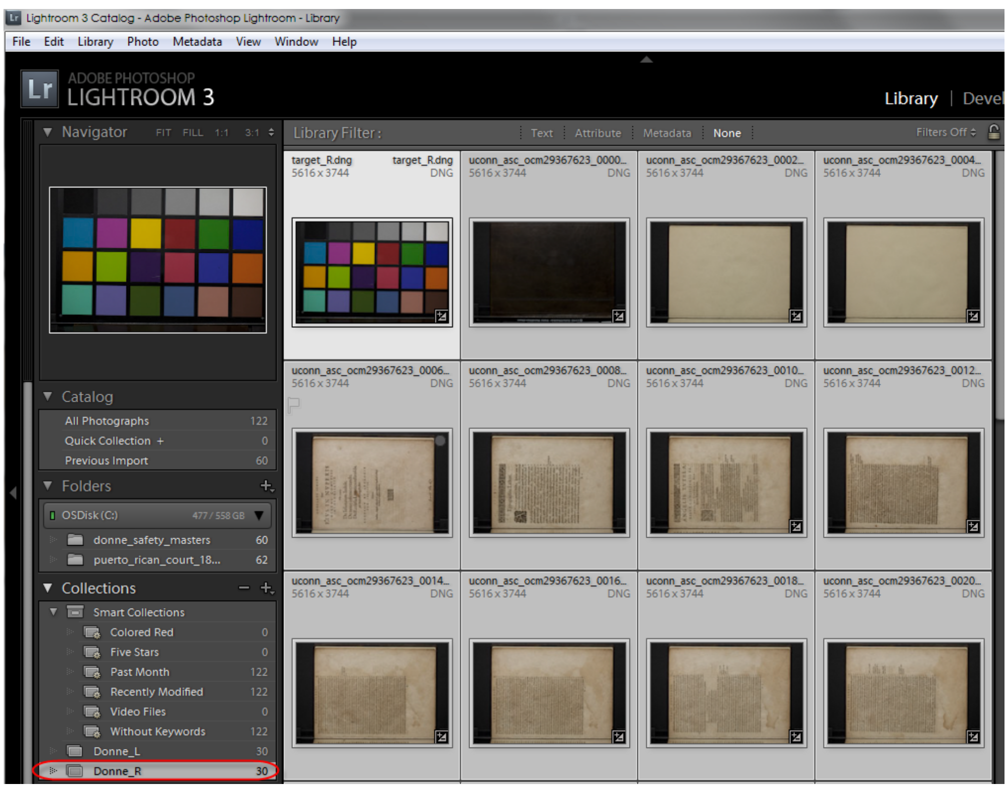


Figures 5. *With all images still selected, a Lightroom Filename text filter for "_L" is applied.*



Figures 6-8. *Filtered images are then added to a new Lightroom collection for editing.*





Figures 9-10. *Similarly, by navigating back to the original safety masters folder all files can then be selected and filtered by "_R" with the results placed in their own "_R" collection, separate from the "_L" pages.*

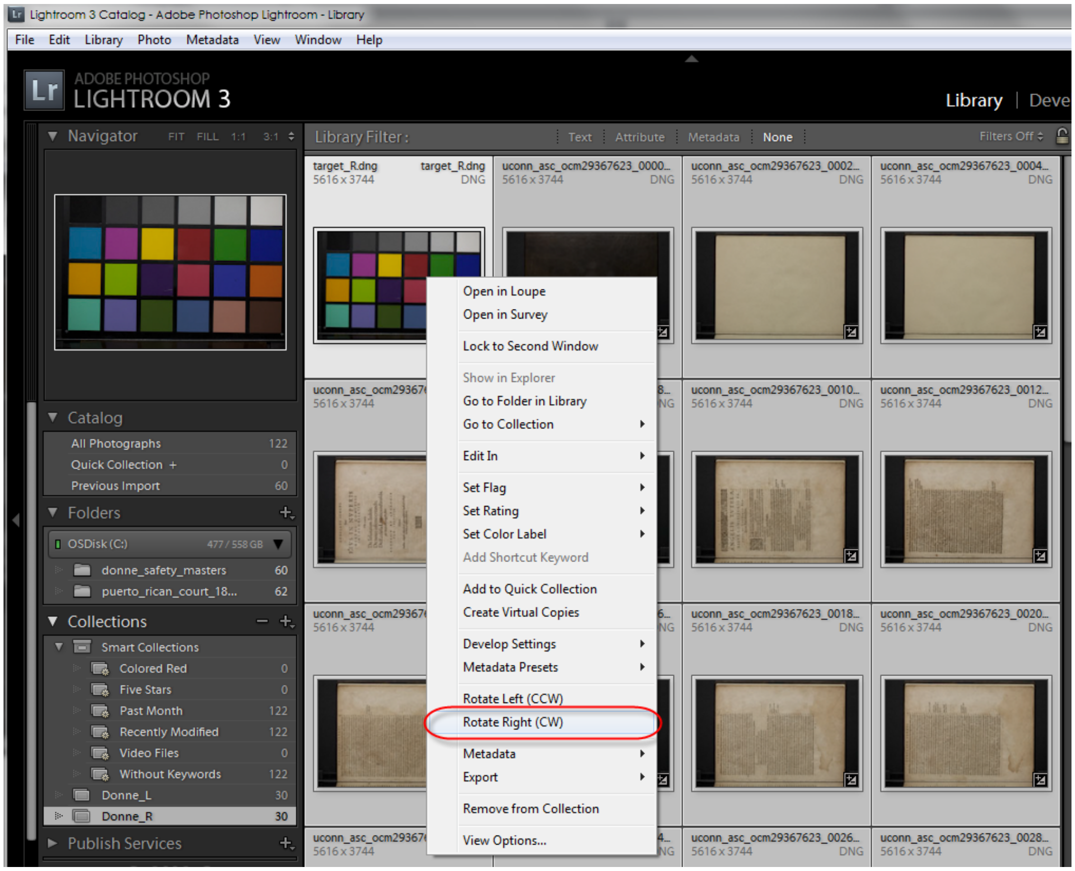


Figure 11. *All images in the Donne_R Collection are rotated right.*

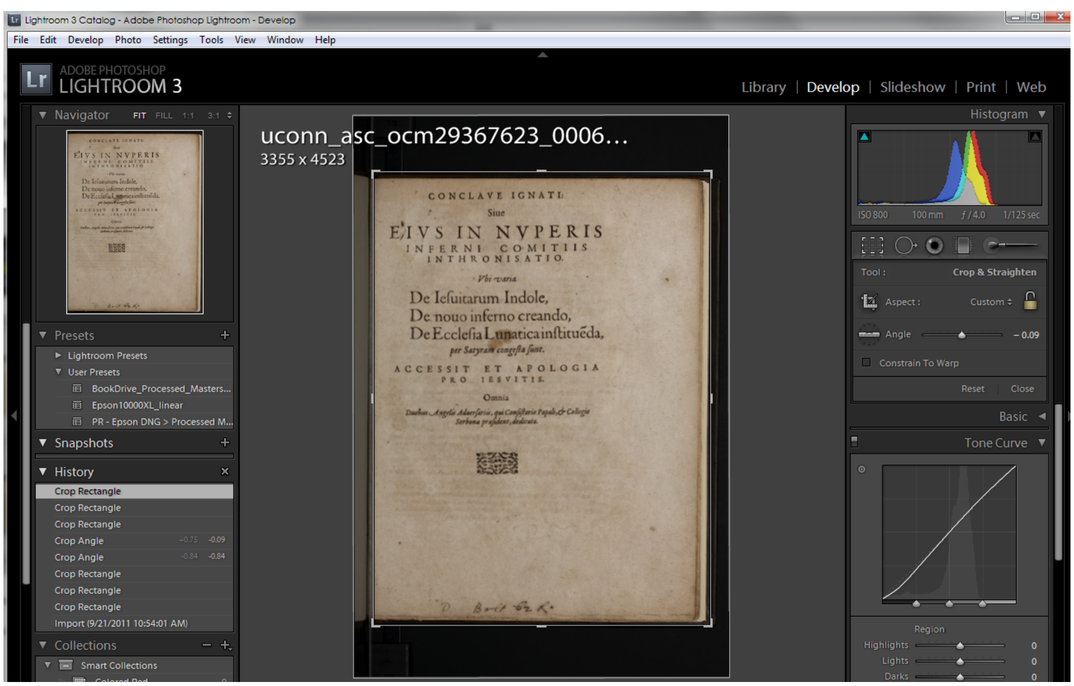


Figure 12. *The first printed page is selected and cropped in Lightroom's Develop module.*

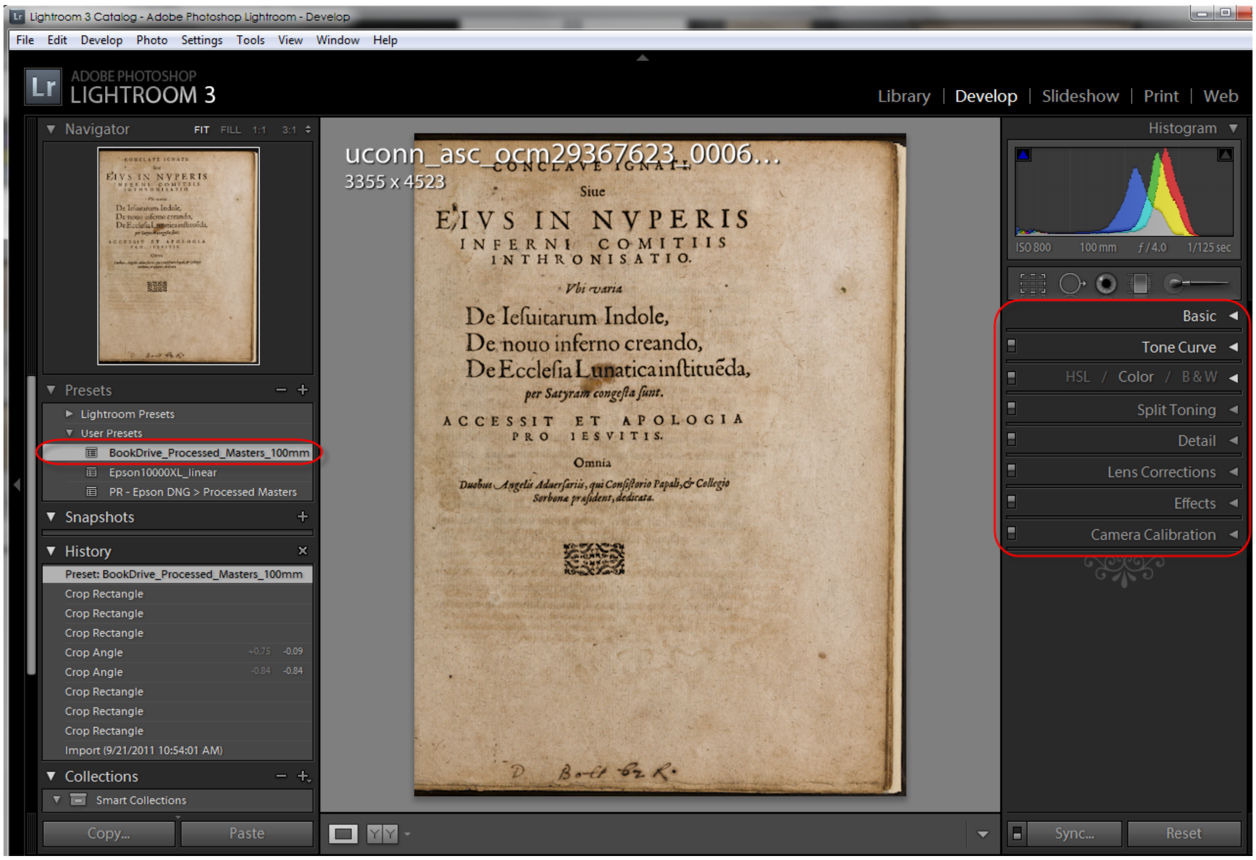
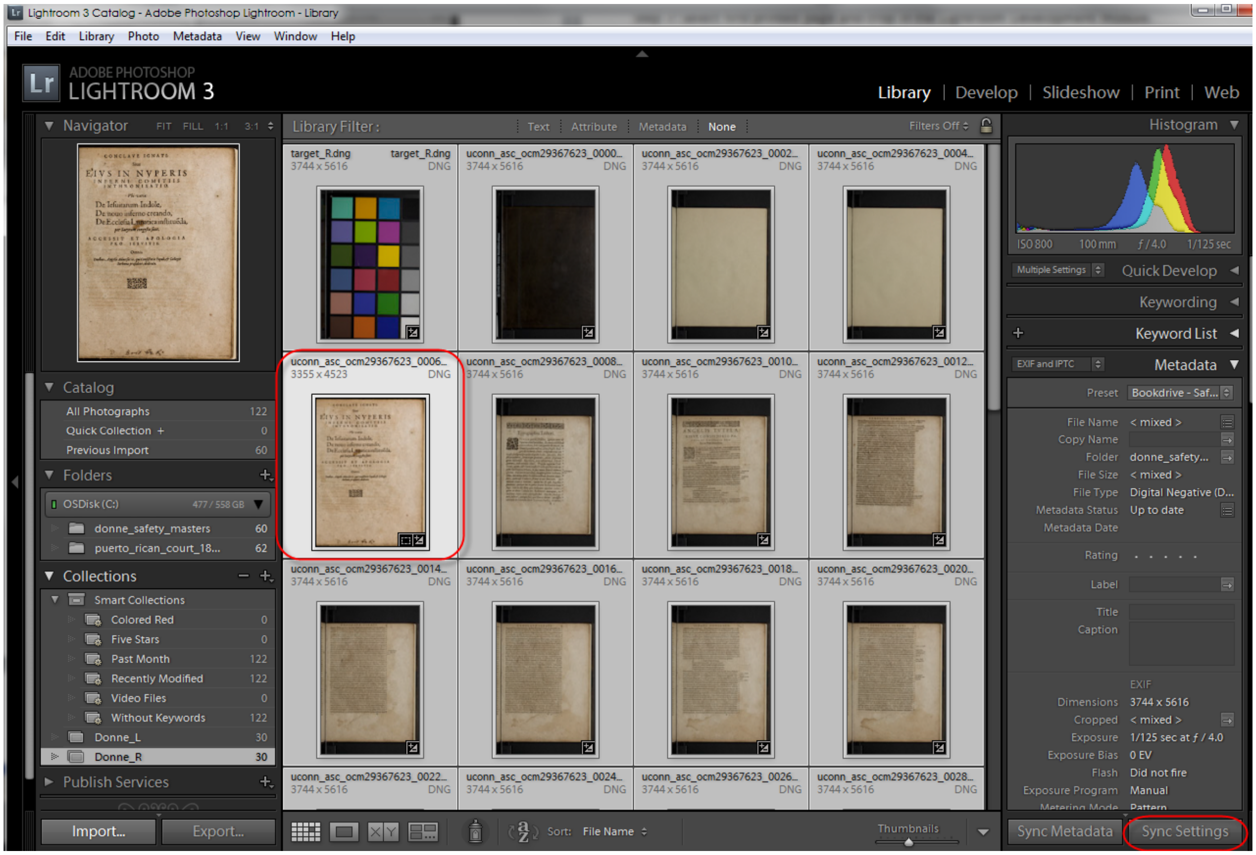
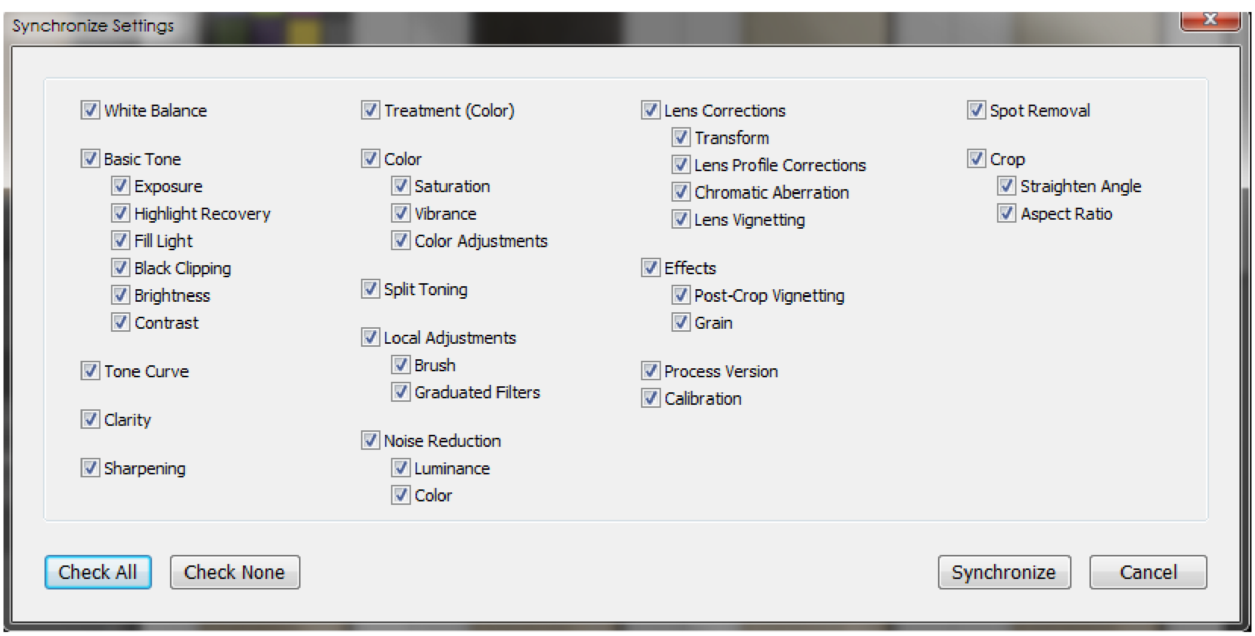
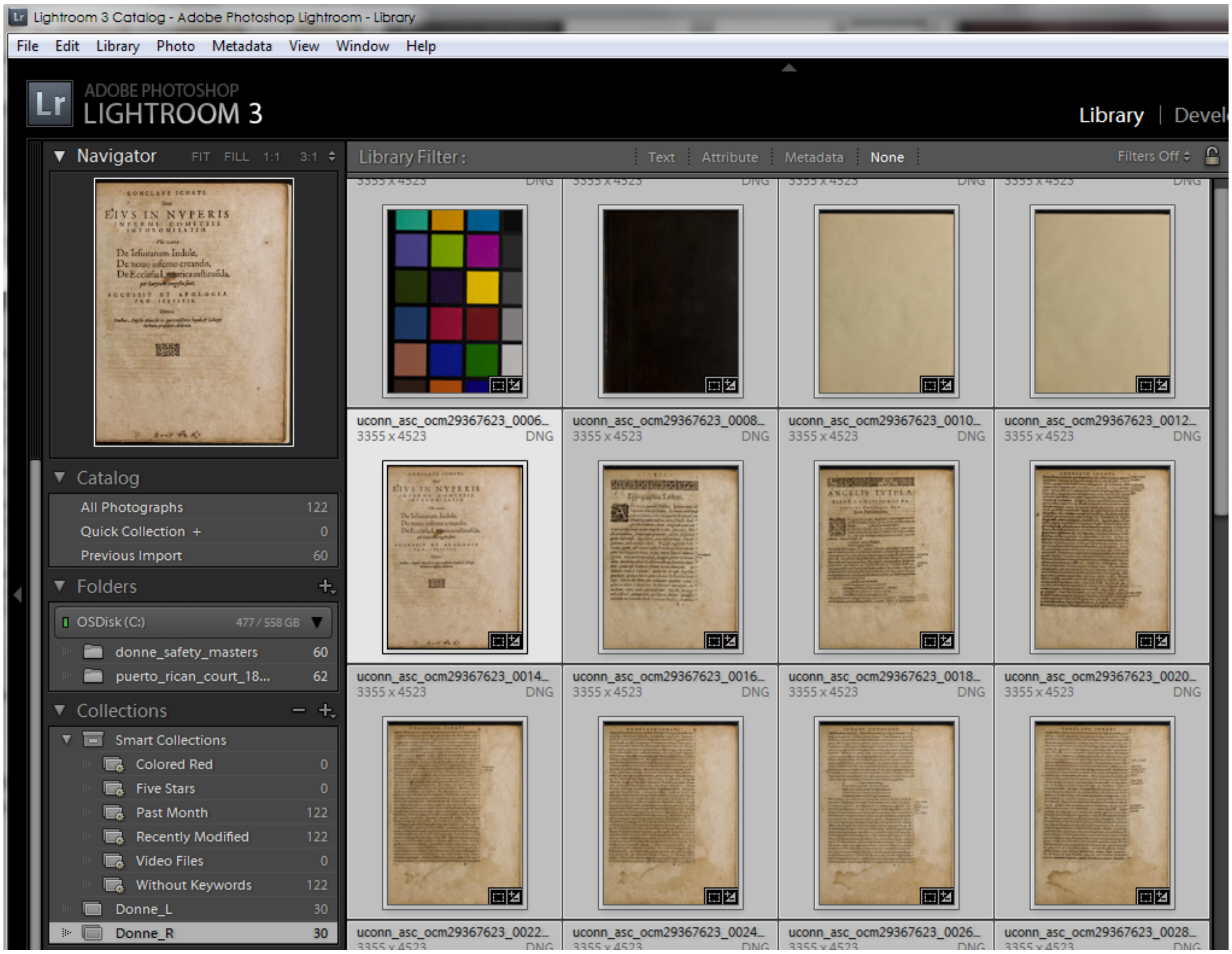


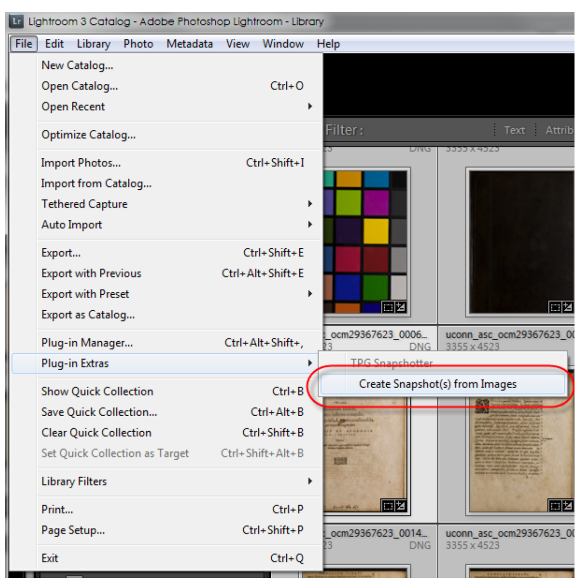
Figure 13. *The page image is further enhanced from either its near linear (zeroed) or Lightroom's shipped default settings state to better meet project reformatting needs as appropriate. Here a previously-created development preset is applied to the image with the main goal of improving text contrast for enhancing downstream OCR success, while also mitigating paper color shifts from such strong tonal adjustment.*

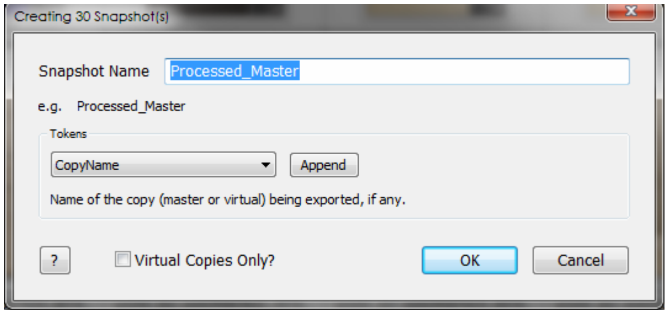

Lightroom 3 Catalog - Adobe Photoshop Lightroom - Library
File Edit Library Photo Metadata View Window Help
ADOBE PHOTOSHOP
LIGHTROOM 3
Library | Develop | Slideshow | Print | Web
Navigator
Library Filter:
Text Attribute Metadata None
Histogram
Catalog
All Photographs 122
Quick Collection + 0
Previous Import 60
Folders
OSDisk (C:)
donne_safety_masters 60
puerto_rican_court_18... 62
Collections
Smart Collections
Colored Red 0
Five Stars 0
Past Month 122
Recently Modified 122
Video Files 0
Without Keywords 122
Donne_L 30
Donne_R 30
Publish Services
Import...
Export...
Quick Develop
Keywording
Keyword List
Metadata
Preset Bookdrive - Saf...
File Name < mixed >
Folder donne_safety...
File Size < mixed >
File Type Digital Negative (D...
Metadata Status Up to date
Rating
Label
Title
Caption
Dimensions 3744 x 5616
Cropped < mixed >
Exposure 1/125 sec at ƒ / 4.0
Exposure Bias 0 EV
Flash Did not fire
Exposure Program Manual
Sort: File Name
Thumbnails
Sync Metadata
Sync Settings

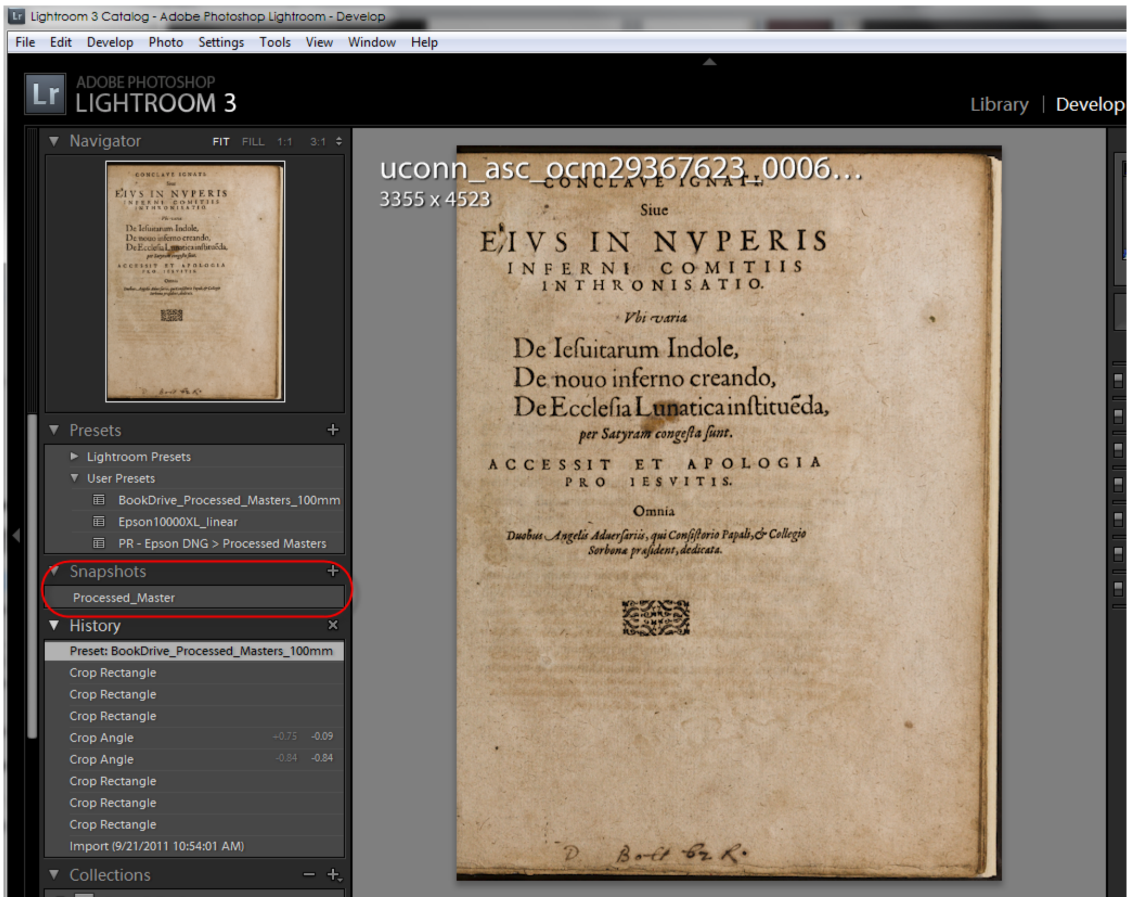

Synchronize Settings
White Balance
Basic Tone
Exposure
Highlight Recovery
Fill Light
Black Clipping
Brightness
Contrast
Tone Curve
Clarity
Sharpening
Treatment (Color)
Color
Saturation
Vibrance
Color Adjustments
Split Toning
Local Adjustments
Brush
Graduated Filters
Noise Reduction
Luminance
Color
Lens Corrections
Transform
Lens Profile Corrections
Chromatic Aberration
Lens Vignetting
Effects
Post-Crop Vignetting
Grain
Process Version
Calibration
Spot Removal
Crop
Straighten Angle
Aspect Ratio
Check All
Check None
Synchronize
Cancel



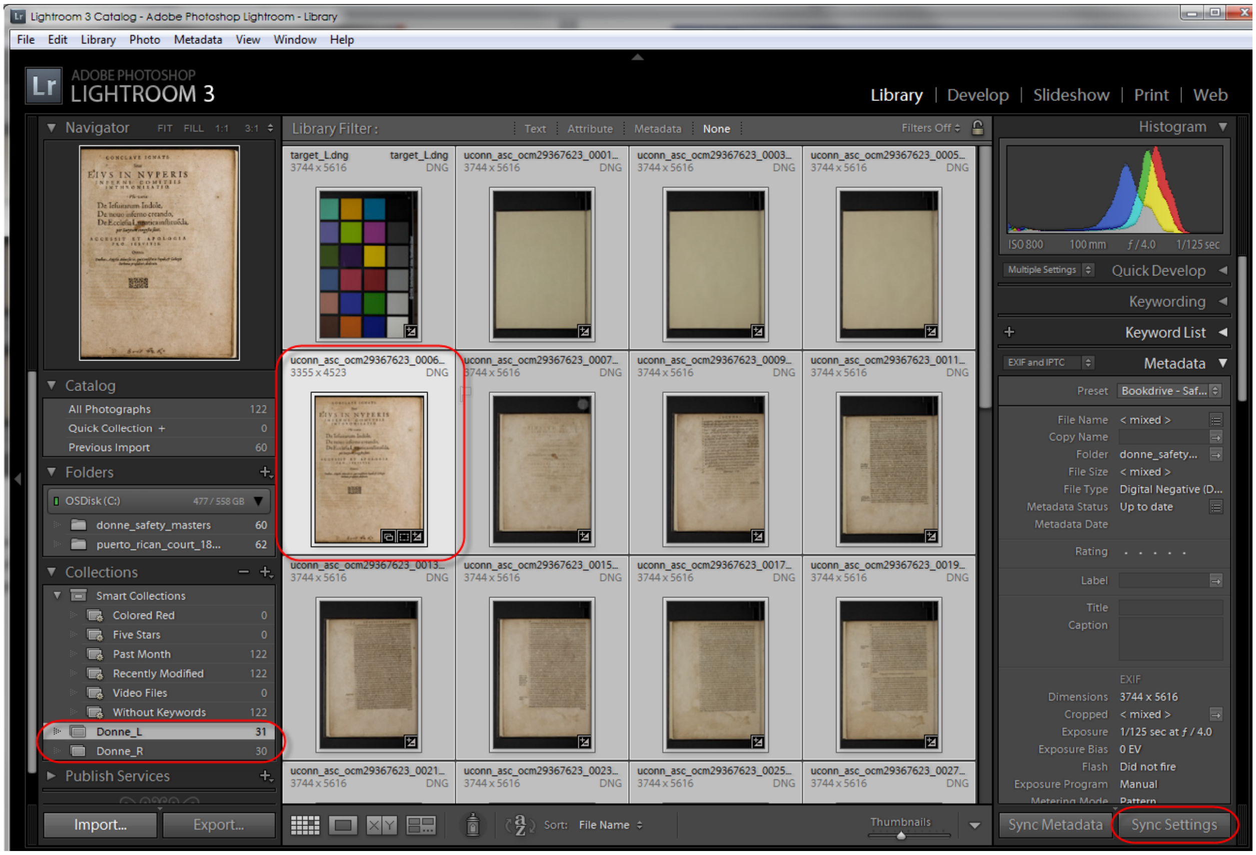
Figures 14-16. *Back in the Library Module, Sync Settings is applied from the just-edited and still active page image to the other selected images in the _R Collection in order to copy both the cropping and development settings just made. Crop boxes are re-aligned (but not resized) on individual images as needed in the Develop Module.*







Figures 17-19. *A "Processed_Master" Snapshot is batch assigned with the Snapshotter plug-in [11] to all edited images in order to better secure and manage particular parametric processing adjustments. Through the application of Snapshots, which can embed such instruction sets into the DNG files themselves, various DNG edited "states" [12][13] can be easily called up in Lightroom (or Adobe Camera Raw). From multiple selected DNGs, Snapshot-controlled "states" can then be quickly batch exported on demand as converted raster formats for various purposes (e.g. "Processed for Text Enhancement," "Processed for Print Reproduction," "Scene Referred State," "Zeroed or Linear Latent State," etc.). See loose manuscript workflow in next section for another example of the use of Snapshots.*



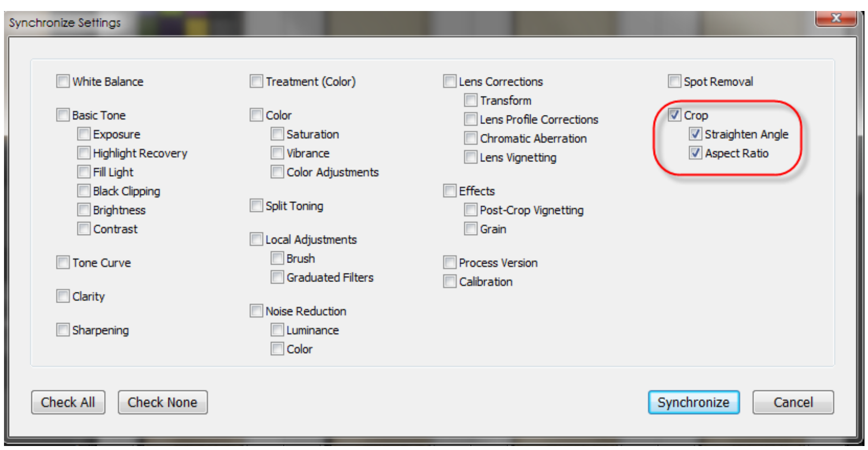


Figure 20-21. *The other _L Collection is navigated to in Library Module, where all images are subsequently selected and rotated. One of the images from the _R Collection is then added to the _L Collection. The Crop setting alone from the added image is synched to the rest of the _L Collection's images. This ensures that page sizing is the same among both right and left hand pages. The _R active image is then removed from the _L Collection. The crop box on one of the new images is re-aligned and then re-synced among just the _L images. The same Develop Preset used previously on the _R Collection images is applied to one image, and then Sync Settings is applied to all images in the _L Collection. A "Processed_Master" Snapshot is added to all _L Collection images with the Snapshotter plugin as described earlier. Finally, still in Lightroom, the safety_master folder is navigated to in order to examine both right and left edited pages together in filename order. Once it has been determined that all images are satisfactory, both the _L and _R Collections are deleted.*

Loose Manuscript Page Workflow: DNG from Scanner Trilinear Array Sensor Data

DNGs can also be created directly from scanners through the use of VueScan software. In this way a measure of parametric editing workflow and image format continuity can be coordinated among a conversion lab's given range of capture devices. As a result, aspects of batch parametric processing need not be completely re-written from scratch for each equipment type but can be re-purposed and shared among a broader spectrum of cameras and scanners.

It bears noting that as opposed to color filter array (CFA) sensor devices like the majority of today's digital cameras, common flatbed scanners employ a trilinear array of RGB-filtered CCD sensor elements [14]. In turn, unlike CFA-based camera DNGs which contain mosaic sensor data, native scanner DNGs are linear encoded RGB files at inception. Such linear (gamma 1.0) DNGs, however, still enjoy many of the same lossless parametric editing efficiencies as camera-based DNGs when manipulated in tools like Lightroom, Adobe Camera Raw, Bibble, etc. In addition, VueScan's default uncompressed DNGs can also be losslessly compressed when subsequently batch processed through such tools or Adobe's DNG Converter. The resulting storage savings of losslessly compressed DNGs (see chart in next section) scale favorably in terms of high volume conversion projects. Also, planned project capture standards may more easily sway towards higher resolution and/or greater bit depth aims since such choices can be less dictated by the elevated storage costs of traditional uncompressed TIFF creation and be more focused on the overall goal of high-quality imaging.

As previously illustrated and in the following demonstration, DNG can be flexibly leveraged across a broad array of project and operational aims. In contrast to proprietary raw specifications, DNG's openly documented architecture uniquely allows the format to be coherently preserved and predictably re-used across platforms and applications. Through the utilization of parametric signposts like "Snapshots," a variety of edited "states" along with various software processing versions can begin to be managed consistently through time.

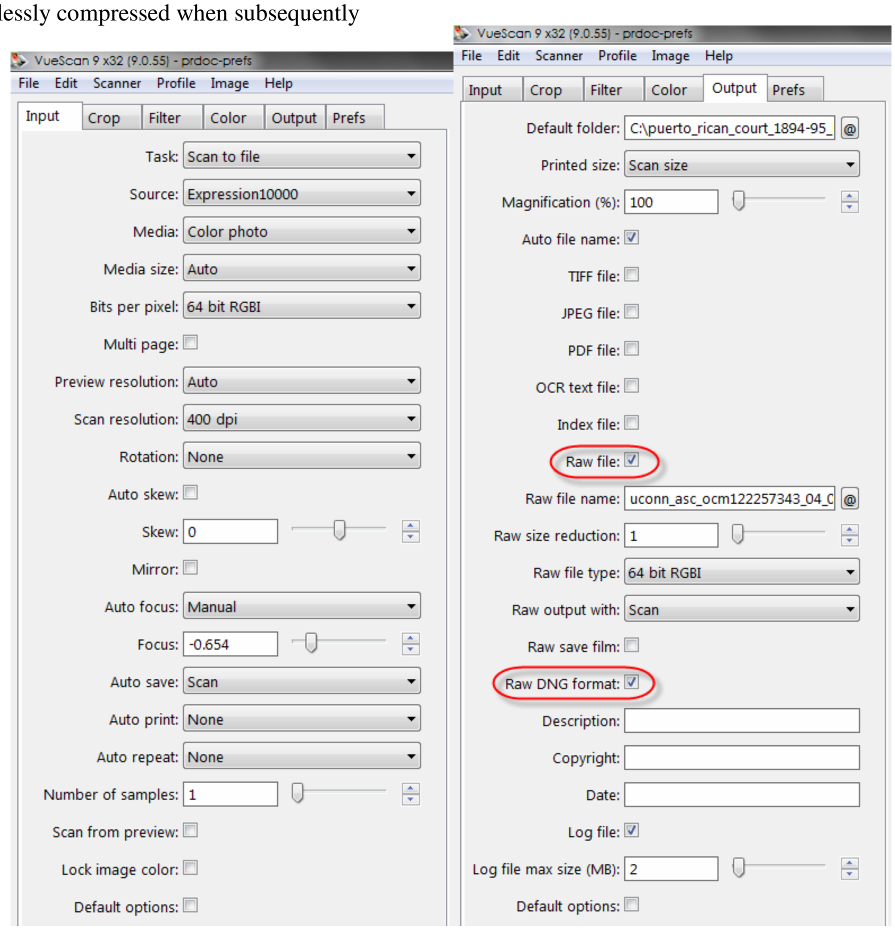


Figure 22. *VueScan v9.0.55 Settings for DNG scanner capture.*

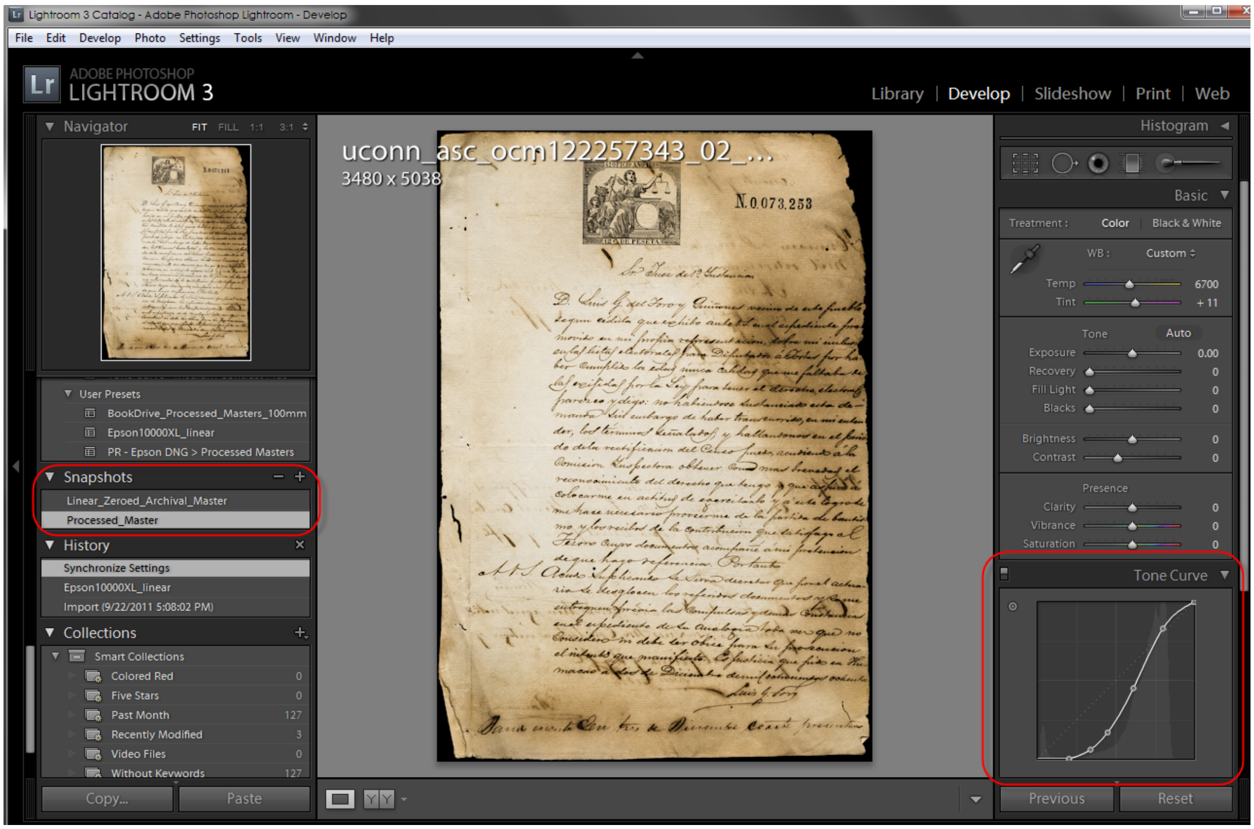


Figure 23. *Saved Snapshots for sample scanner DNG from a 19th century Puerto Rican civil court manuscript reformatting project. In this example, the Processed_Master Snapshot is activated and shows the steep parametric tone curve applied to the manuscript to better enhance front-side handwritten legibility from backside handwritten bleed-through. Additional parametric Color adjustments include both negative Hue and negative Saturation to Yellow that are used in order to better manage resulting paper color shifts that result from previous tone curve handwriting enhancements.*

Lossless JPEG 2000 as Raster Archival Master File Format Alternative to TIFF

One of the simpler ways to begin to explore the advantages of JPEG 2000 is to consider its losslessly compressed use as an archival raster format substitute to uncompressed TIFF. On average, a given lossless encoded JPEG 2000 file will be 1/3 the size of the same image saved as uncompressed TIFF all without loss of any image information. When factored into a given institution's total number of archival image files, substantial, scalable data storage savings can be readily achieved.

Lossless JPEG 2000 files can be batch-created directly from camera raw files or converted DNGs in the following automated manner.

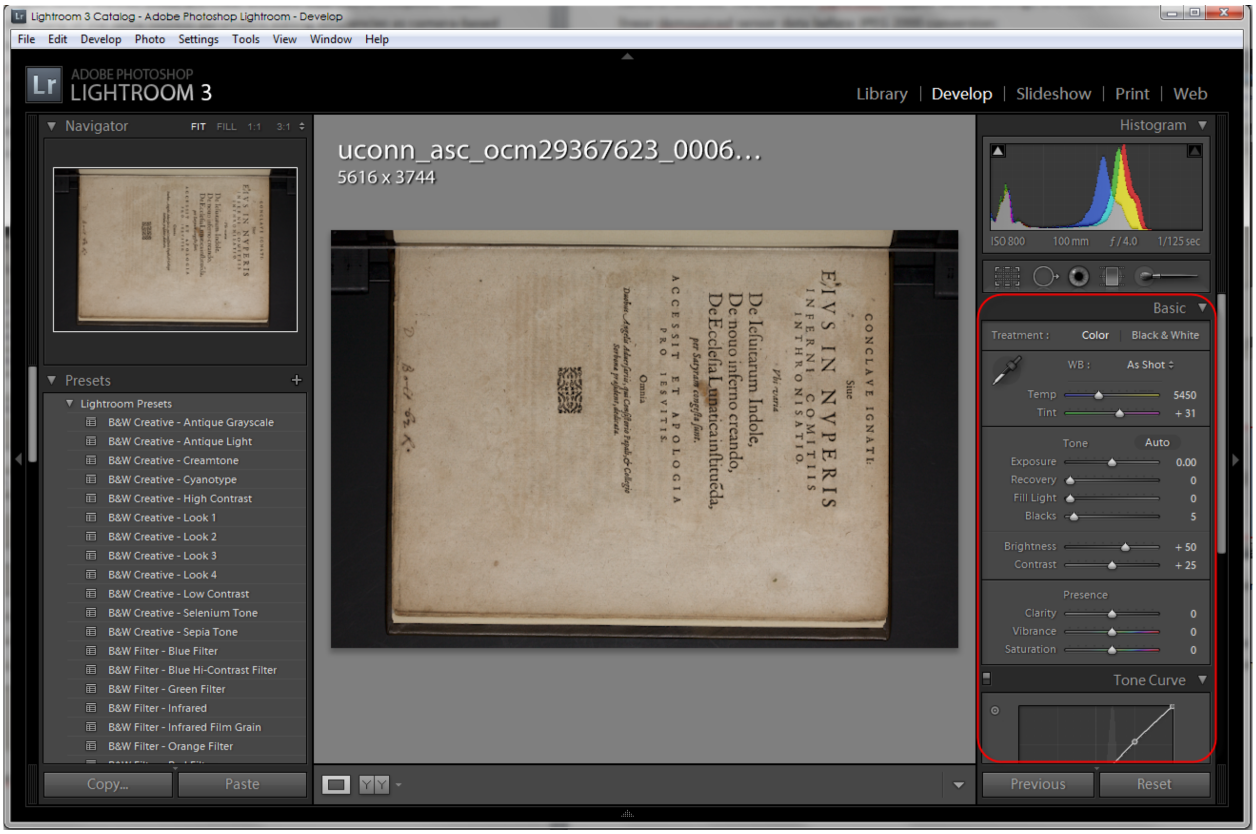


Figure 24. *A determination of what level of processing, if any, is made for the raw files prior to JPEG 2000 conversion. Here, Lightroom's shipped default presets are shown which employ gamma correction to the near linear demosaiced sensor data and can be synched to all monograph raw images prior to JPEG 2000 conversion. For more specific scene-referred JP2000 rendering, more fine-tuned presets can be created, stored, and likewise synched.*

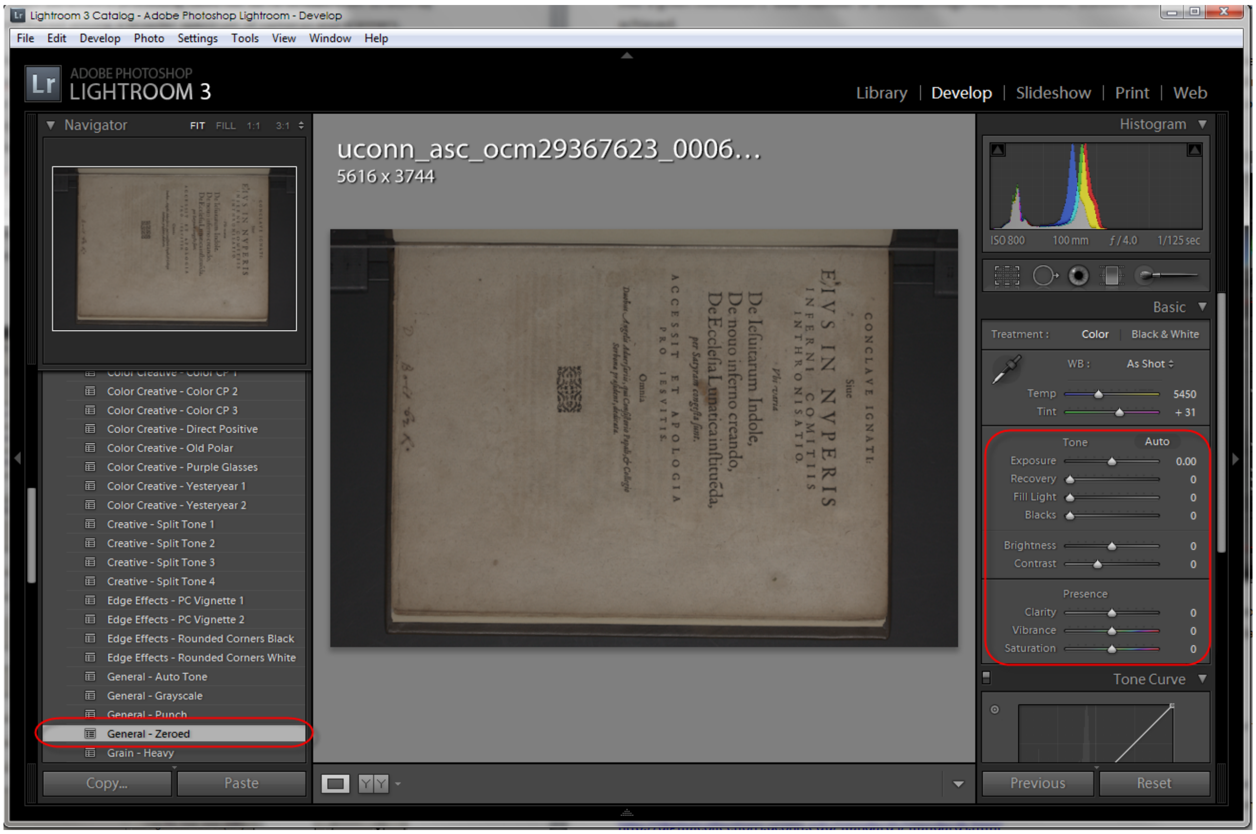


Figure 25. *Or all raw images can be "zeroed" before JPEG 2000 conversion which in essence leaves the raw files in a near latent, linear state.*

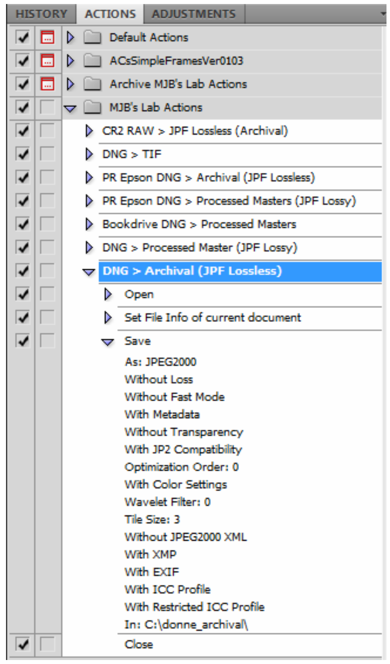


Figure 26. *A Photoshop action for DNG > lossless JPEG 2000 conversion is created. (Note: source and destination folders used while creating the action are irrelevant. See next steps.)*

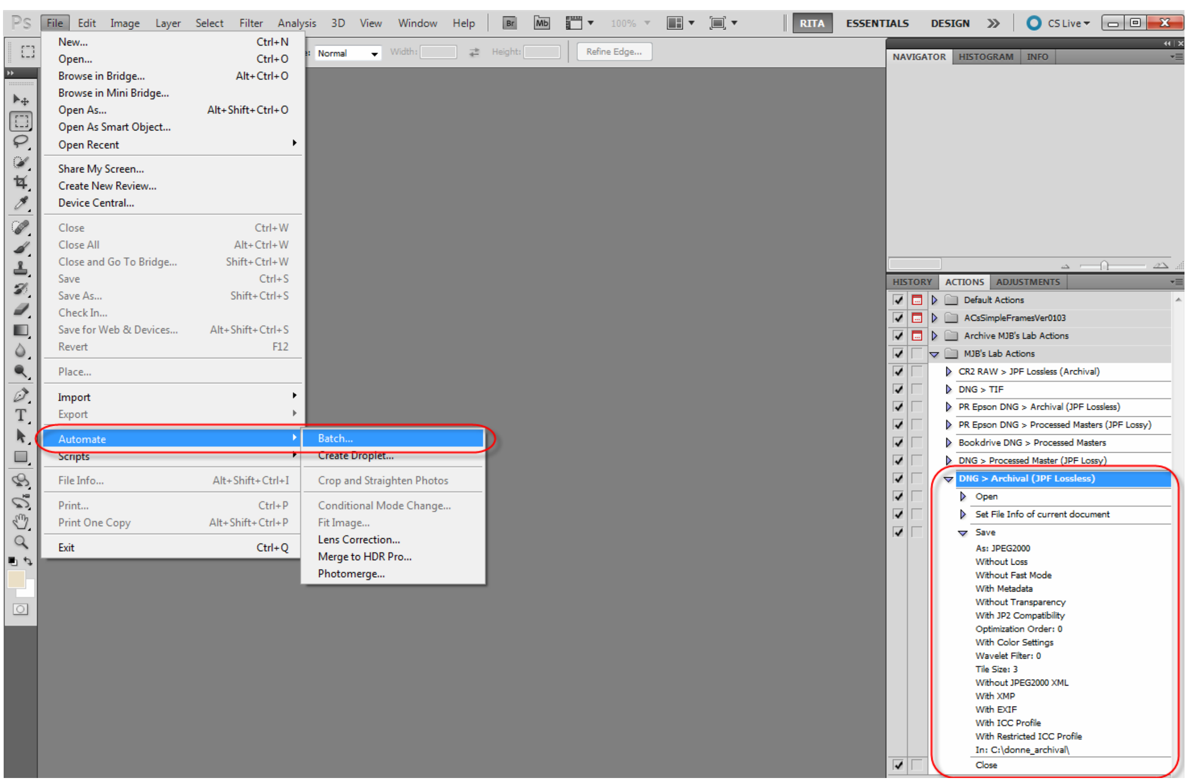


Figure 27. *A destination folder is created for the JPEG 2000 archival files that are about to be encoded (e.g. book_title_archival). Photoshop is then opened. The "DNG > Archival (JPF Lossless)" action is chosen. File > Automate > Batch. DNG > Archival (JPF Lossless) action should be pre-selected.*

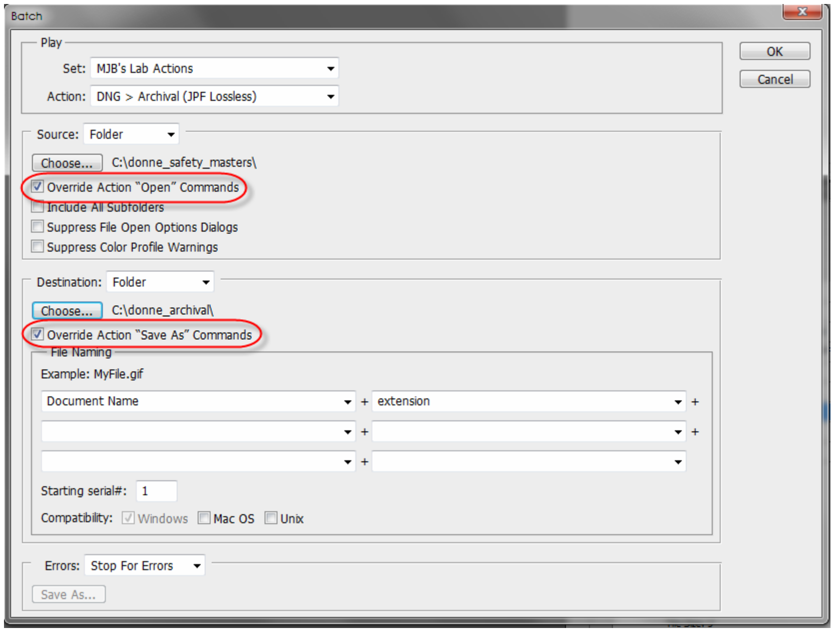


Figure 28. *The folder, c:\book_title_safety_masters\ (e.g. c:\donne_safety_masters) is chosen as the source folder with the "Override Action Open Commands" ticked. The archival folder made earlier on C:\ is selected as the destination folder with "Override Action Save As Commands" ticked [15]. Lossless JPEG 2000 files then can be batch processed from the raw source files.*

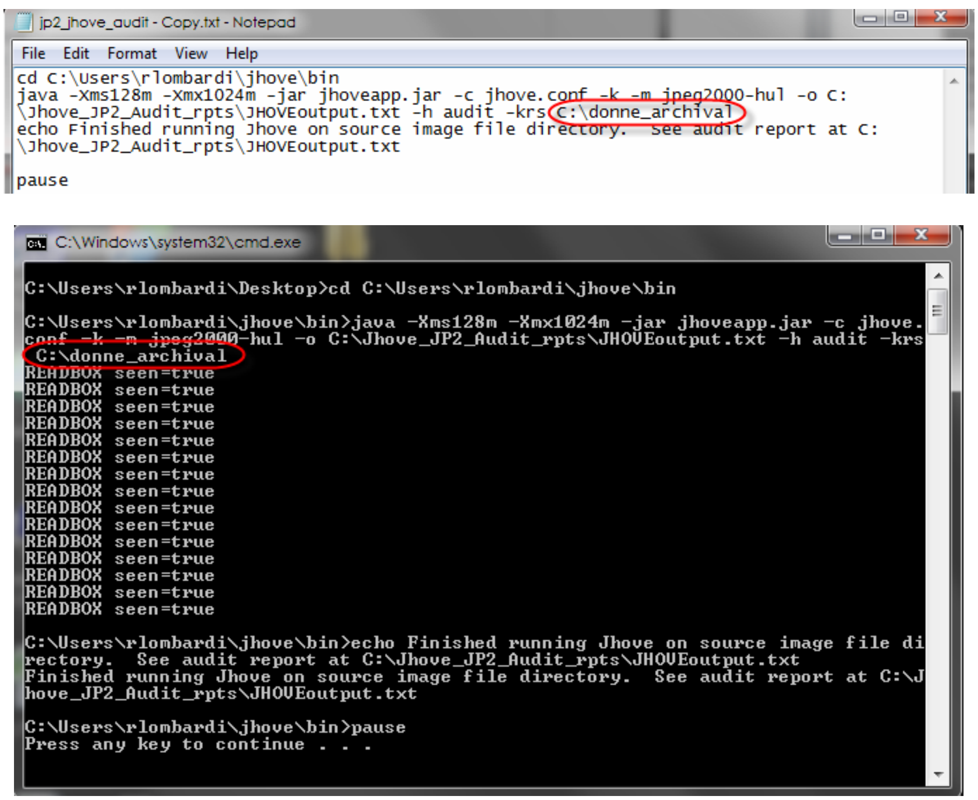
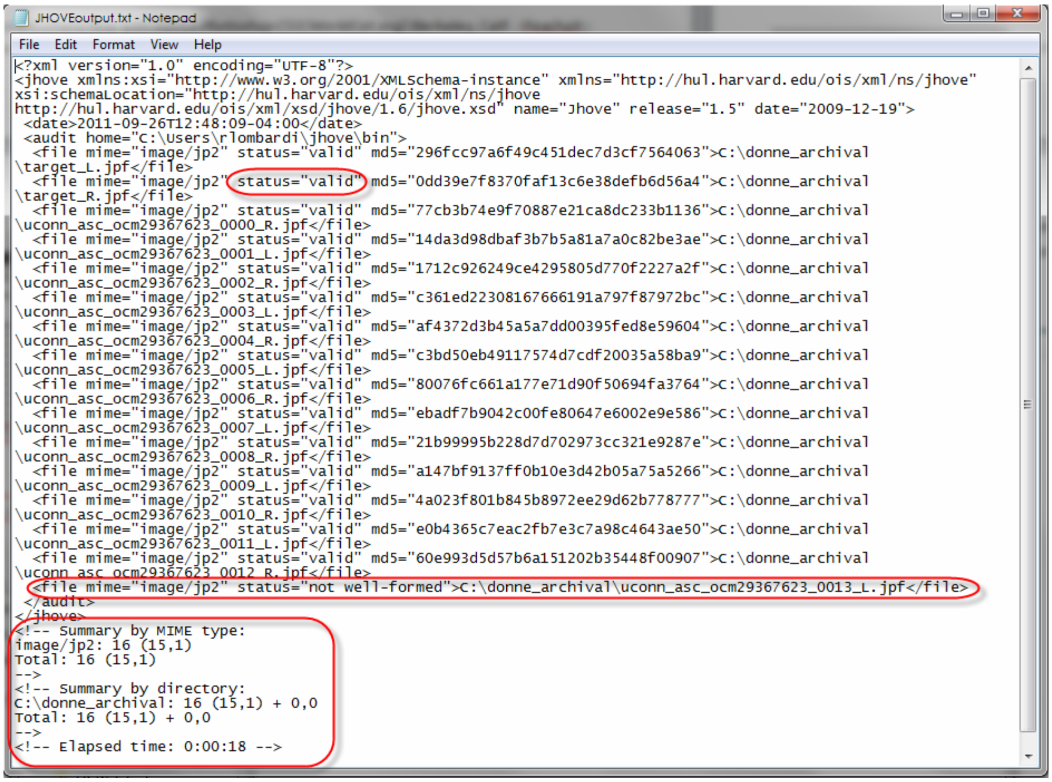


Figure 29-30. *A Jhove v1.5 audit is run on resulting .jpf archival files to check for any encoding errors. UConn employs the simple batch script [16] illustrated above to run Jhove from the command line on the archival folder's files. The script also instructs Jhove to output audit results into C:\Jhove_JP2_Audit_rpts\JHOVEoutput.txt. To run the script: at the desktop where it is normally saved for ease of use by imaging technicians, the file extension of jp2_jhove_audit.cmd is temporarily changed to jp2_jhove_audit.txt and the file is opened in Notepad. The end of the 3rd line of batch script is then edited to reflect the folder name of the given monograph's images, (i.e. c:\book_title_archival). After edit, jhove_audit.txt is saved. The file's extension is then manually changed back to .cmd on the desktop. The jp2_jhove_audit.cmd batch file can then be double-clicked from desktop to start the script that audits the files in c:\book_title_archival for problems. (Note: Jhove run on a 16 file sub-sample of larger c:\donne_archival folder in figure above to better illustrate script messaging).*



JHOVEoutput.txt - Notepad

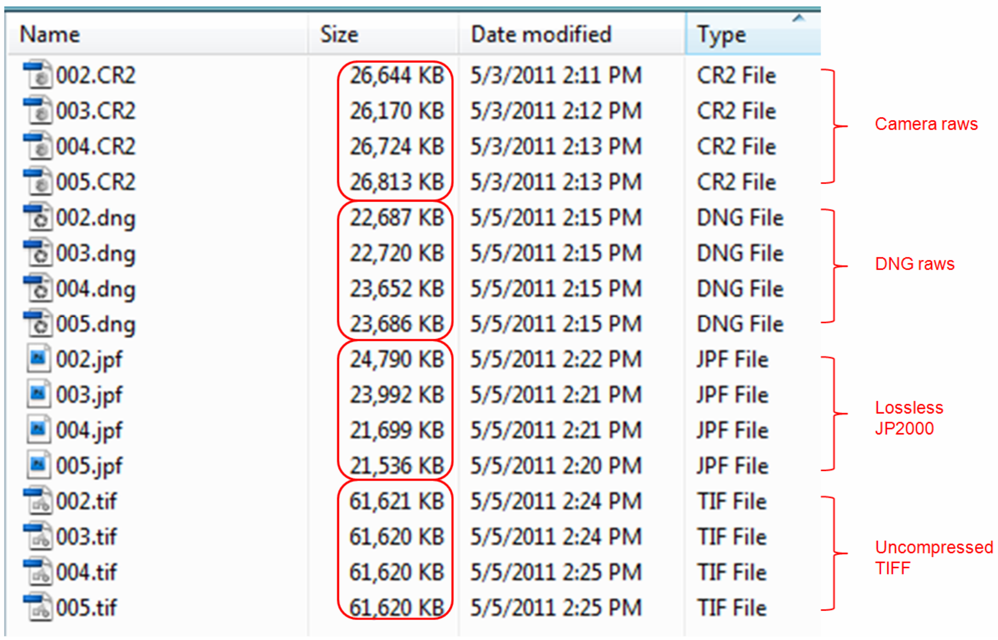
File Edit Format View Help


```
<?xml version="1.0" encoding="UTF-8"?>
<jhove xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://hul.harvard.edu/ois/xml/ns/jhove"
xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/jhove
http://hul.harvard.edu/ois/xml/xsd/jhove/1.6/jhove.xsd" name="Jhove" release="1.5" date="2009-12-19">
 <date>2011-09-26T12:48:09-04:00</date>
 <audit home="C:\Users\rlombardi\jhove\bin">
  <file mime="image/jp2" status="valid" md5="296fcc97a6f49c451dec7d3cf7564063">C:\donne_archival
\target_L.jpf</file>
  <file mime="image/jp2" status="valid" md5="0dd39e7f8370faf13c6e38defb6d56a4">C:\donne_archival
\target_R.jpf</file>
  <file mime="image/jp2" status="valid" md5="77cb3b74e9f70887e21ca8dc233b1136">C:\donne_archival
\uconn_asc_ocm29367623_0000_R.jpf</file>
  <file mime="image/jp2" status="valid" md5="14da3d98dbaf3b7b5a81a7a0c82be3ae">C:\donne_archival
\uconn_asc_ocm29367623_0001_L.jpf</file>
  <file mime="image/jp2" status="valid" md5="1712c926249ce4295805d770f2227a2f">C:\donne_archival
\uconn_asc_ocm29367623_0002_R.jpf</file>
  <file mime="image/jp2" status="valid" md5="c361ed22308167666191a797f87972bc">C:\donne_archival
\uconn_asc_ocm29367623_0003_L.jpf</file>
  <file mime="image/jp2" status="valid" md5="af4372d3b45a5a7dd00395fed8e59604">C:\donne_archival
\uconn_asc_ocm29367623_0004_R.jpf</file>
  <file mime="image/jp2" status="valid" md5="c3bd50eb49117574d7cdf20035a58ba9">C:\donne_archival
\uconn_asc_ocm29367623_0005_L.jpf</file>
  <file mime="image/jp2" status="valid" md5="80076fc661a177e71d90f50694fa3764">C:\donne_archival
\uconn_asc_ocm29367623_0006_R.jpf</file>
  <file mime="image/jp2" status="valid" md5="ebadf7b9042c00fe80647e6002e9e586">C:\donne_archival
\uconn_asc_ocm29367623_0007_L.jpf</file>
  <file mime="image/jp2" status="valid" md5="21b99995b228d7d702973cc321e9287e">C:\donne_archival
\uconn_asc_ocm29367623_0008_R.jpf</file>
  <file mime="image/jp2" status="valid" md5="a147bf9137ff0b10e3d42b05a75a5266">C:\donne_archival
\uconn_asc_ocm29367623_0009_L.jpf</file>
  <file mime="image/jp2" status="valid" md5="4a023f801b845b8972ee29d62b778777">C:\donne_archival
\uconn_asc_ocm29367623_0010_R.jpf</file>
  <file mime="image/jp2" status="valid" md5="e0b4365c7eac2fb7e3c7a98c4643ae50">C:\donne_archival
\uconn_asc_ocm29367623_0011_L.jpf</file>
  <file mime="image/jp2" status="valid" md5="60e993d5d57b6a151202b35448f00907">C:\donne_archival
\uconn_asc_ocm29367623_0012_R.jpf</file>
  <file mime="image/jp2" status="not well-formed">C:\donne_archival\uconn_asc_ocm29367623_0013_L.jpf</file>
 </audit>
</jhove>
<!-- Summary by MIME type:
image/jp2: 16 (15,1)
Total: 16 (15,1)
-->
<!-- Summary by directory:
C:\donne_archival: 16 (15,1) + 0,0
Total: 16 (15,1) + 0,0
-->
<!-- Elapsed time: 0:00:18 -->
```


Figure 31. *From the resulting C:\Jhove_JP2_Audit_rpts folder, the .txt file inside is then opened to view errors. "Not well-formed" is an error, "valid" is a good file. The numbers at the bottom of the report indicate that out of 16 files, 15 were good and 1 was bad. Any bad files are re-encoded, and then the audit steps are repeated once again until no errors occur. (Note: Bad file was manually corrupted in figure above prior to audit checking for illustration purposes). The c:\book_title_archival folder is finally copied to archival storage, and then deleted from C:\ drive.*

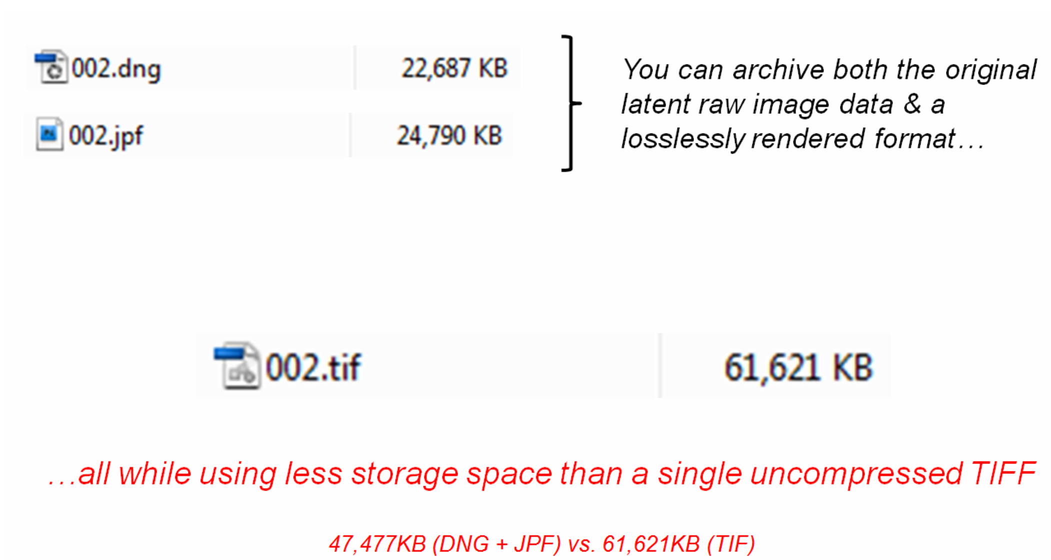
The following illustration summarizes some of the scalable storage advantages of archiving both lossless JPEG 2000 [17] and raw DNGs for a given camera image vs. uncompressed TIF. By taking advantage of the lossless compression efficiencies of DNG and JPEG 2000, institutions not willing at this point in time to only save raw files can still reap the robust data preservation and processing gains of raw while maintaining the traditional benefits of rendered still image archiving. Significantly, this can all be achieved while taking up less storage space than a single uncompressed, rendered TIF.

If…



Name	Size	Date modified	Type
002.CR2	26,644 KB	5/3/2011 2:11 PM	CR2 File
003.CR2	26,170 KB	5/3/2011 2:12 PM	CR2 File
004.CR2	26,724 KB	5/3/2011 2:13 PM	CR2 File
005.CR2	26,813 KB	5/3/2011 2:13 PM	CR2 File
002.dng	22,687 KB	5/5/2011 2:15 PM	DNG File
003.dng	22,720 KB	5/5/2011 2:15 PM	DNG File
004.dng	23,652 KB	5/5/2011 2:15 PM	DNG File
005.dng	23,686 KB	5/5/2011 2:15 PM	DNG File
002.jpf	24,790 KB	5/5/2011 2:22 PM	JPF File
003.jpf	23,992 KB	5/5/2011 2:21 PM	JPF File
004.jpf	21,699 KB	5/5/2011 2:21 PM	JPF File
005.jpf	21,536 KB	5/5/2011 2:20 PM	JPF File
002.tif	61,621 KB	5/5/2011 2:24 PM	TIF File
003.tif	61,620 KB	5/5/2011 2:24 PM	TIF File
004.tif	61,620 KB	5/5/2011 2:25 PM	TIF File
005.tif	61,620 KB	5/5/2011 2:25 PM	TIF File

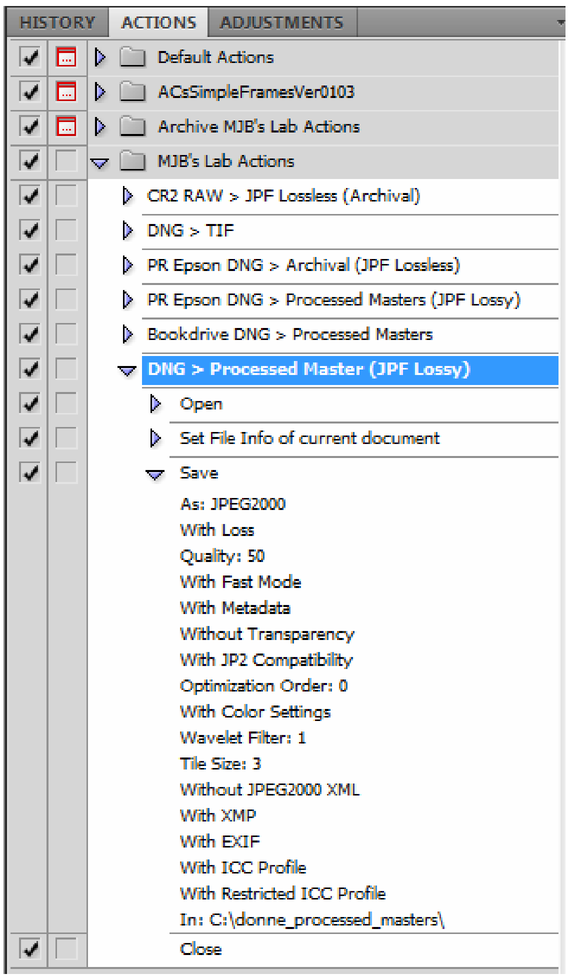
Then…



Lossy JPEG 2000 Processed Master File Format

Through collaboration with software engineer, Hank Bromley, from the Internet Archive (IA) the author has tailored the UConn lab's monograph and manuscripts workflows to integrate with IA's batch ingest protocols. This has allowed the UConn Libraries' lab to function much like an IA scan center for online delivery of these material types. Part of this process is the creation of lossy (but visually lossless) JPEG 2000 processed master files, grouped into .tar files, one "tarball" of all page images per monograph volume. Lossy, irreversible JPEG 2000 is chosen because of its possible visually lossless compression and highly efficient storage savings which scale favorably across all aspects of the combined workflow (i.e. tarball upload, local and IA archiving, automated IA OCR, IA eBook format encodings, and interactive online "bookreader" interface generation). An example of the final results for one volume may be viewed at http://www.archive.org/details/conclaveignati00donn.

DNG Safety Masters with "Processed_Master" Snapshots are the source for such rendered JPEG 2000 processed master images. The DNG Snapshots normally represent the source images parametrically rotated, cropped, with applied tonal adjustments best suited for high OCR success as described earlier. Lossy, but visually lossless, JPEG 2000s are then batch created along with embedded technical metadata through Photoshop from the DNGs in the following way.



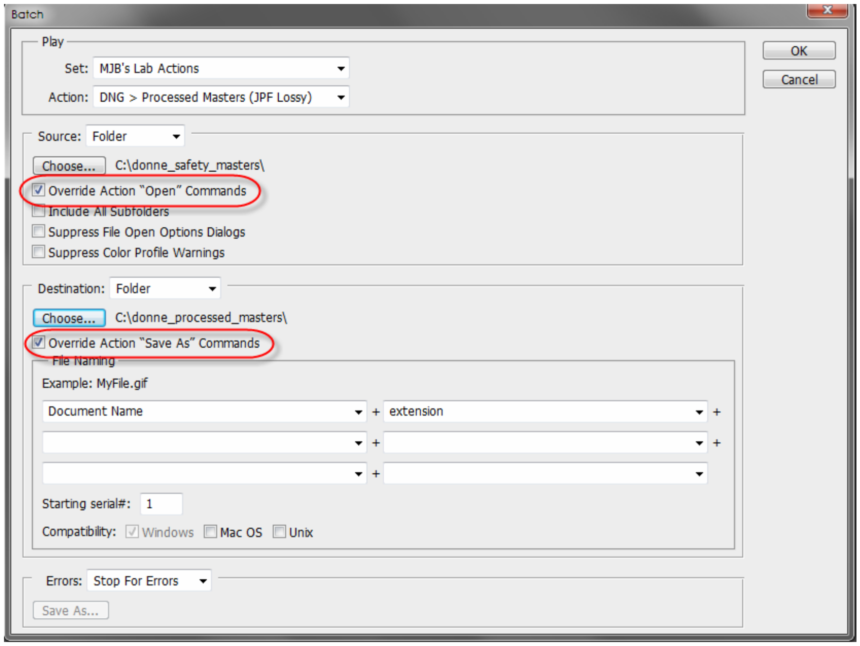


Figure 32-33. *In Photoshop, "DNG > Processed Master (JPF Lossy)" action is selected. File > Automate > Batch is navigated to in order to apply above action and create JPEG 2000 processed masters from DNG files' "Processed_Master" Snapshots. A Jhove Audit on the resulting new JPEG 2000 processed masters is then run to check for encoding errors. Any bad files are then re-encoded, and the audit process is repeated until no errors occur.*

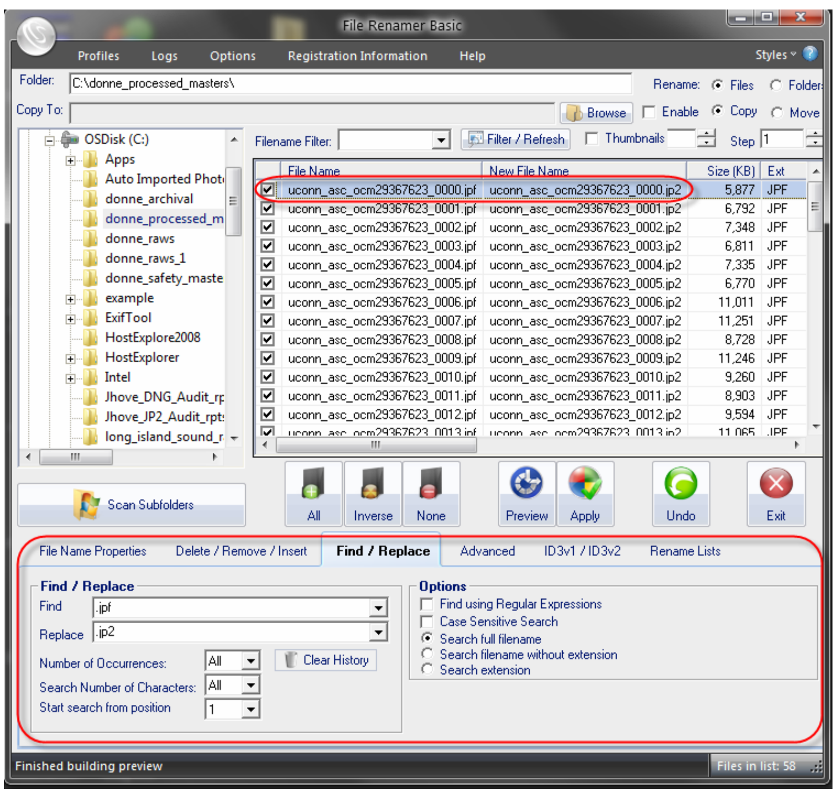


Figure 34. *Files are renamed with jp2 extension for broader ease of use. Note: this step is possible because original .jpf files are batch encoded with "JP2 compatibility" (see Photoshop action save step). The c:\book_title_processed_masters folder is copied to archival storage. c:\book_title_processed_masters is then deleted.*

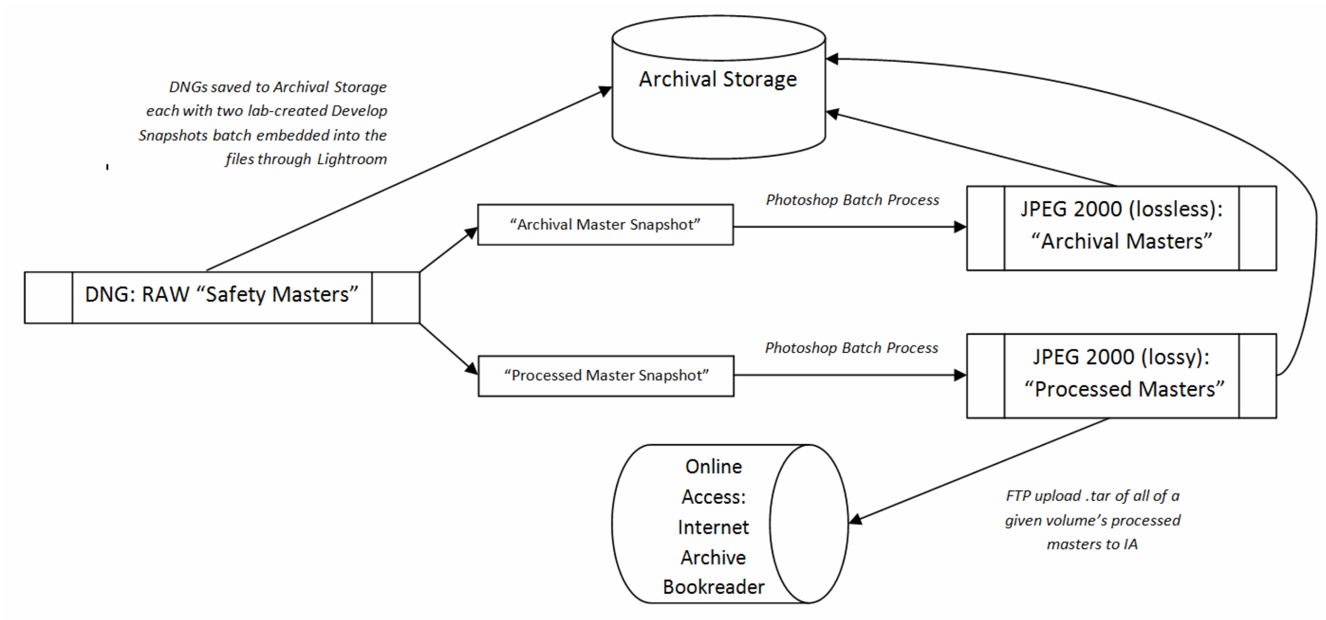


Figure 35. *Schematic View of UConn Monograph Conversion Workflow*

Leveraging Embedded Process Metadata in XMP

File-embedded XMP and its support for IPTC Core opens up new opportunities to create more robust still image files [18][19][20]. Such files can contain not only device-generated Exif information and parametric editing instruction tags (including Snapshots), but can also contain IPTC Core elements that can be edited either individually in Photoshop or in batches through Lightroom metadata presets and/or Adobe Bridge/Photoshop metadata templates.

The advantages of such additional embedded descriptive metadata are many. Individual still image files can be less dependent upon traditional external catalogs for their descriptions and can in essence be self-describing assets with sufficient descriptive information. This is of particular interest as images are exported and re-purposed beyond the institutional gates of their creation and become de-coupled from their original hosted settings.

Important file creation information or "process metadata" can also be efficiently embedded to include details of technical provenance and image editing [21]. Such particulars can greatly assist in future large-scale migrations and/or accurate file replications as hardware, workstation OS, and post-processing software versions change through time.

Finally, once embedded in all files, both descriptive and technical process metadata greatly assist in original digital asset management (DAM) system imports and/or future DAM platform migrations. As the vast majority of DAMs move toward fuller XMP compliance, catalog database migrations and their inherent problems may be made easier with more fully self-described source files that in essence become their own best record. Additionally, XMP is serialized in XML and stored using a subset of the W3C Resource Description Framework (RDF) [22]. As such, XMP's structure incorporates well when repurposed and leveraged through OAIS digital preservation technology stacks like Archivematica and repository frameworks such as Fedora.

What follows are examples of how the UConn Libraries' lab has begun to embed and standardize such metadata into the various still image files examined throughout this study.

METADATA	
▼ File Properties	
Filename	uconn_asc_ocm29367623_0006_R.dng
Preserved Filename	
Document Type	DNG image
Application	Adobe Photoshop Lightroom 3.4.1
Date Created	9/6/2011, 1:29:01 PM
Date File Modified	Today, 2:29:51 PM
File Size	22.55 MB
Dimensions	5616 x 3744
Dimensions (in inches)	
Resolution	
Bit Depth	16
Color Mode	RGB
Color Profile	Untagged
Notes	
Supports XMP	Yes
▼ IPTC Core	
Creator	Rita Lombardi
Creator: Job Title	Digital Photographer
Creator: Address	University of Connecticut, Homer Babbidge Library
Creator: City	Storrs
Creator: State/Province	Connecticut
Creator: Postal Code	06269-205
Creator: Country	USA
Creator: Phone(s)	
Creator: Email(s)	digitalcollections@uconn.edu
Creator: Website(s)	http://digitalcollections.uconn.edu/
Headline	
Description	
Keywords	
IPTC Subject Code	
Description Writer	
Date Created	9/6/2011, 1:29:01 PM
Intellectual Genre	
IPTC Scene Code	
Sublocation	
City	
State/Province	
Country	
ISO Country Code	
Title	Conclave Ignati : siue eivs in nvperis inferni comitiis in...
Job Identifier	
Instructions	Original RAW .cr2 image capture: Canon 5D Mark II digital SLR still camera mounted on ATIZ BookDrive DIY book cradle with ATIZ BookDrive Capture v4.2.5.0. RAW .cr2 images then DNG-converted and parametrically edited in Lightroom v3.4.1.
Credit Line	
Source	
Copyright Notice	
Copyright Status	Unknown

Rights Usage Terms	The image is owned, held, or licensed by the University of Connecticut and is available for personal, non-commercial, and educational use, provided that ownership of the image is properly cited. Any commercial use of the image, without the written permission of the University of Connecticut, is strictly prohibited.
▼ Fonts	
▼ Linked Files	
▼ Plates	
▼ Document Swatches	
▼ Camera Data (Exif)	
Exposure	1/125 s at f/4.0
Exposure Bias Value	
Exposure Mode	Manual
Exposure Index	
Exposure Program	Manual
Brightness Value	
ISO Speed Ratings	800
Focal Length	100.0 mm
Focal Length in 35mm Film	
Lens	EF100mm f/2.8 Macro USM
Max Aperture Value	f/2.8
Artist	Rita Lombardi
Date Time	9/6/2011, 1:36:27 PM
Date Time Original	9/6/2011, 1:29:01 PM
Date Time Digitized	9/6/2011, 1:29:01 PM
Flash	Did not fire, compulsory mode
Flash Energy	
Metering Mode	Evaluative
Orientation	Rotate 180°
Exif Color Space	
Light Source	
Subject Distance	1.0 m
User Comment	
Subject Area	
Custom Rendered	Normal Process
White Balance	Manual
Digital Zoom Ratio	
Scene Capture Type	Standard
Gain Control	
Contrast	
Saturation	
Sharpness	
Subject Distance Range	
Image Unique ID	
Sensing Method	
Image Description	
Subject Location	
File Source	
Make	Canon
Model	Canon EOS 5D Mark II
Serial Number	1821104842

Figure 36. *Sample XMP snippet from DNG Safety Master. Note: Additional metadata written to file through Lightroom metadata preset. Develop settings including Lightroom-created "Snapshots" not shown in figure.*

METADATA	
▼ File Properties	
Filename	uconn_asc_ocm29367623_0005_L.jpf
Preserved Filename	
Document Type	jpf
Application	Adobe Photoshop CS5.1 Windows
Date Created	9/6/2011, 10:24:50 AM
Date File Modified	Today, 11:23:33 AM
File Size	16.63 MB
Dimensions	5616 x 3744
Dimensions (in inches)	14.0" x 9.4"
Resolution	400 ppi
Bit Depth	
Color Mode	RGB
Color Profile	Adobe RGB (1998)
Notes	
Supports XMP	Read-Only
▼ IPTC Core	
Creator	Rita Lombardi
Creator: Job Title	Digital Photographer
Creator: Address	University of Connecticut, Homer Babbidge Library
Creator: City	Storrs
Creator: State/Province	Connecticut
Creator: Postal Code	06269-205
Creator: Country	USA
Creator: Phone(s)	
Creator: Email(s)	digitalcollections@uconn.edu
Creator: Website(s)	http://digitalcollections.uconn.edu/
Headline	
Description	
Keywords	
IPTC Subject Code	
Description Writer	
Date Created	9/6/2011, 10:24:50 AM
Intellectual Genre	
IPTC Scene Code	
Sublocation	
City	
State/Province	
Country	
ISO Country Code	
Title	Conclave Ignati : siue eivs in nvperis inferni comitiis in thr...
Job Identifier	
Instructions	Original RAW .cr2 image capture: Canon 5D Mark II digital SLR still camera mounted on ATIZ BookDrive DIY book cradle with ATIZ BookDrive Capture v4.2.5.0. RAW .cr2 images then DNG-converted and parametrically edited in Lightroom v3.4.1. DNG converted to lossless JPEG 2000 with Adobe Photoshop CS5 v12.1 batch process: MJB Lab action, 400 ppi JPEG 2000 (.jpf) without Loss, Without Fast Mode, With Metadata, Without Transparency, With JP2 Compatibility, Optimization Order Growing Thumbnail, Download Rate 56.6 Kbps, With Color Settings, Compliance General Device, Wavelet Filter Integer, Tile Size 1024x1024, Without JPEG 2000 XML, With XMP, With EXIF, With ICC Profile, With Restricted ICC Profile.

Copyright Notice	
Copyright Status	Unknown
Rights Usage Terms	The image is owned, held, or licensed by the University of Connecticut and is available for personal, non-commercial, and educational use, provided that ownership of the image is properly cited. Any commercial use of the image, without the written permission of the University of Connecticut, is strictly prohibited.
▼ Fonts	
▼ Linked Files	
▼ Plates	
▼ Document Swatches	
▼ Camera Data (Exif)	
Exposure	1/125 s at f/4.0
Exposure Bias Value	
Exposure Mode	Manual
Exposure Index	
Exposure Program	Manual
Brightness Value	
ISO Speed Ratings	800
Focal Length	100.0 mm
Focal Length in 35mm Film	
Lens	EF100mm f/2.8 Macro USM
Max Aperture Value	f/2.8
Artist	Rita Lombardi
Date Time	Today, 11:22:49 AM
Date Time Original	9/6/2011, 10:24:50 AM
Date Time Digitized	9/6/2011, 10:24:50 AM
Flash	Did not fire, compulsory mode
Flash Energy	
Metering Mode	Evaluative
Orientation	Normal
Exif Color Space	Uncalibrated
Light Source	
Subject Distance	1.0 m
User Comment	
Subject Area	
Custom Rendered	Normal Process
White Balance	Manual
Digital Zoom Ratio	
Scene Capture Type	Standard
Gain Control	
Contrast	
Saturation	
Sharpness	
Subject Distance Range	
Image Unique ID	
Sensing Method	
Image Description	
Subject Location	
File Source	
Make	Canon
Model	Canon EOS 5D Mark II
Serial Number	1921202185

Figure 37. *Sample XMP snippet from Archival File (Lossless JPEG 2000): Note "Instructions" field used for technical metadata describing post-processing and JP2000 "save as" profile. Metadata written to file by Photoshop Action step. Information in remaining fields carried over from safety master source file. All XMP incorporated into JPEG 2000 UUID (Universally Unique Identifier) box.*

METADATA	
▼ File Properties	
Filename	uconn_asc_ocm29367623_0006.jpf
Preserved Filename	
Document Type	jpf
Application	Adobe Photoshop CS5.1 Windows
Date Created	9/6/2011, 1:29:01 PM
Date File Modified	Today, 11:29:30 AM
File Size	5.71 MB
Dimensions	3559 x 4650
Dimensions (in inches)	8.9" x 11.6"
Resolution	400 ppi
Bit Depth	
Color Mode	RGB
Color Profile	sRGB IEC61966-2.1
Notes	
Supports XMP	Read-Only
▼ IPTC Core	
Creator	Rita Lombardi
Creator: Job Title	Digital Photographer
Creator: Address	University of Connecticut, Homer Babbidge Library
Creator: City	Storrs
Creator: State/Province	Connecticut
Creator: Postal Code	06269-205
Creator: Country	USA
Creator: Phone(s)	
Creator: Email(s)	digitalcollections@uconn.edu
Creator: Website(s)	http://digitalcollections.uconn.edu/
Headline	
Description	
Keywords	
IPTC Subject Code	
Description Writer	
Date Created	9/6/2011, 1:29:01 PM
Intellectual Genre	
IPTC Scene Code	
Sublocation	
City	
State/Province	
Country	
ISO Country Code	
Title	Conclave Ignati : siue eivs in nvperis inferni comitiis in thr…
Job Identifier	
Instructions	Original RAW .cr2 image capture: Canon 5D Mark II digital SLR still camera mounted on ATIZ BookDrive DIY book cradle with ATIZ BookDrive Capture v4.2.5.0. RAW .cr2 images then DNG-converted and parametrically edited in Lightroom v3.4.1. Opened edited DNG files using Adobe Camera Raw v6.4.1 in Photoshop CS5 v12.1 and saved as lossy jpeg 2000 files: save as .jpf with Loss, Quality 50, Without Fast Mode, With Metadata, Without Transparency, With JP2 Compatibility, Optimization Order Growing Thumbnail, Download Rate 56.6 Kbps, With Color Settings, Compliance General Device, Wavelet Filter Integer, Tile Size 1024x1024, Without JPEG 2000 XML, With XMP, With EXIF, With ICC Profile, With Restricted ICC Profile. File extension renamed from .jpf to .jp2 with FileRenamer 6.3.0.

Copyright Notice	
Copyright Status	Unknown
Rights Usage Terms	The image is owned, held, or licensed by the University of Connecticut and is available for personal, non-commercial, and educational use, provided that ownership of the image is properly cited. Any commercial use of the image, without the written permission of the University of Connecticut, is strictly prohibited.
▼ Fonts	
▼ Linked Files	
▼ Plates	
▼ Document Swatches	
▼ Camera Data (Exif)	
Exposure	1/125 s at f/4.0
Exposure Bias Value	
Exposure Mode	Manual
Exposure Index	
Exposure Program	Manual
Brightness Value	
ISO Speed Ratings	800
Focal Length	100.0 mm
Focal Length in 35mm Film	
Lens	EF100mm f/2.8 Macro USM
Max Aperture Value	f/2.8
Artist	Rita Lombardi
Date Time	Today, 11:29:19 AM
Date Time Original	9/6/2011, 1:29:01 PM
Date Time Digitized	9/6/2011, 1:29:01 PM
Flash	Did not fire, compulsory mode
Flash Energy	
Metering Mode	Evaluative
Orientation	Normal
Exif Color Space	sRGB
Light Source	
Subject Distance	1.0 m
User Comment	
Subject Area	
Custom Rendered	Normal Process
White Balance	Manual
Digital Zoom Ratio	
Scene Capture Type	Standard
Gain Control	
Contrast	
Saturation	
Sharpness	
Subject Distance Range	
Image Unique ID	
Sensing Method	
Image Description	
Subject Location	
File Source	
Make	Canon
Model	Canon EOS 5D Mark II
Serial Number	1821104842

Figure 38. *Sample XMP snippet from Processed Master File (Lossy JPEG 2000): Note "Instructions" field used for technical metadata describing post-processing and JPEG 2000 "save as" profile. Metadata written to file by Photoshop Action step. Information in remaining fields carried over from safety master source file. All XMP incorporated into JPEG 2000 UUID box.*

Conclusion

Today, recent developments in digital reformatting have included a growing movement toward making such conversions more broadly operational, larger scale, and systemic [23][24][25]. Simultaneously, as the software and formats that surround still imaging evolve, a greater need for more robust and flexible digital objects is becoming apparent to meet novel repurposing needs [26][27]. In turn, decisions with regard to the scalable use of raw still image file archiving and processing, and data compression in general are important to consider when both quantity and quality are concurrent goals in today's reformatting ecosystem. Preserving the expertise of trained digital imaging technicians and the full sensitivities of the enlarging array of capture devices that they operate must be done now more than ever in both an efficient and extensible way to meet the requirements of feasible operational growth, new digital object use, and well managed storage over time. In so doing, institutions can more fully preserve and further utilize the fruits of their substantial investments in both digital conversion staff and equipment.

References

[1] Bennett, Michael J., and F. Barry Wheeler. “Raw as Archival Still Image Format: A Consideration.” The Society for Imaging Science and Technology Archiving 2010 Final Program and Proceedings (2010): 185-193.

[2] Krogh, Peter. Non-Destructive Imaging: An Evolution of Rendering Technology, n.d. http://www.adobe.com/digitalimag/pdfs/non_destructive_imaging.pdf.

[3] Anderson, Richard. “Raw vs. Rendered | dpBestflow.” dpBestflow, n.d. http://dpbestflow.org/node/264.

[4] Fraser, Bruce. “Understanding Digital Raw Capture”, n.d. http://www.adobe.com/products/photoshop/pdfs/understanding_digitalrawcapture.pdf.

[5] Russotti, Patti, and Richard Anderson. Digital photography best practices and workflow handbook: a guide to staying ahead of the workflow curve. Burlington, MA: Focal Press, 2010.

[6] Krogh, Peter. The DAM book: digital asset management for photographers. 2nd ed. Beijing; Cambridge [Mass.]: O’Reilly, 2009.

[7] Allen, Elizabeth, and Sophie Triantaphillidou. The manual of photography. 10th ed. Oxford; Burlington, MA: Elsevier/Focal Press, 2011.

[8] Fraser, Bruce, and Jeff Schewe. Real world Camera Raw with Adobe Photoshop CS3. Berkeley, Calif.: Peachpit; London: Pearson Education [distributor], 2008.

[9] Krogh, Peter. Non-Destructive Imaging: An Evolution of Rendering Technology, n.d. http://www.adobe.com/digitalimag/pdfs/non_destructive_imaging.pdf.

[10] “Digital Collections | University of Connecticut Libraries: Standards”, September 27, 2011. http://digitalcollections.uconn.edu/standards/standards.html.

[11] Friedl, Jeffrey. “Snapshotter Plugin | The Photo Geek.” Snapshotter Plugin, The Photo Geek, n.d. http://thephotogeek.com/lightroom/snapshotter/.

[12] Arms, Caroline A., Carl Fleischhauer, and Jimi Jones. “Quality and Functionality Factors for Still Images”, n.d. http://www.digitalpreservation.gov/formats/content/still_quality.shtml.

[13] “ISO 22028-1:2004 - Photography and graphic technology -- Extended colour encodings for digital image storage, manipulation and interchange -- Part 1: Architecture and requirements”, 2004. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37161.

[14] Peres, Michael R. The Focal encyclopedia of photography: digital imaging, theory and applications, history, and science. 4th ed. Amsterdam; Boston: Elsevier/Focal Press, 2007.

[15] Fraser, Bruce, and Jeff Schewe. Real world Camera Raw with Adobe Photoshop CS3. Berkeley, Calif.: Peachpit; London: Pearson Education [distributor], 2008.

[16] Bennett, Michael J. Jhove Audit Batch File, 2010.

[17] Lowe, David, and Michael J. Bennett. “A Status Report on JPEG 2000 Implementation for Still Images: The UConn Survey.” The Society of Imaging Science and Technology Archiving 2009, Final Program and Proceedings (2009): 209-212.

[18] Adobe.“XMP Specification Part 3, Storage in Files. July 2010.” http://www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/XMPSpecificationPart3.pdf.

[19] Christensen, S. O., & Dunlop, D. “The Case for Implementing Core Descriptive Embedded Metadata at the Smithsonian.” Proceedings of the International Conference on Dublin Core and Metadata Applications. (2010)

[20] Christensen, S. O., Dunlop, D., Pilsk, S., Snyder, R., Nguyen, D., Stauderman, S., Smith, S., et al. “Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images.” (2010) http://hdl.handle.net/10088/9719

[21] Metadata Working Group. “Guidlines for Handling Image Metadata, Version 2.0.” http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf.

[22] TC 130. “ISO 16684-1:2012 - Graphic Technology -- Extensible Metadata Platform (XMP) Specification -- Part 1: Data Model, Serialization and Core Properties.” http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57421.

[23] Erway, Ricky, and Jennifer Schaffner. Shifting Gears: Gearing Up to Get Into the Flow. Dublin, OH: OCLC, 2007. http://www.oclc.org/research/publications/library/2007/2007-02.pdf

[24] Smith, Stanley, Alan Newman, Chris Gallagher, Chris Edwards, and John ffrench. “Speed the Plow: Rapid Capture Digital Workflow Handout”. Portland, OR, 2009. http://www.mcn.edu/conference/mcn2009/Smith_Getty_Rapid_Capture_Project_Photographs.pdf.

[25] Erway, Ricky. Rapid Capture: Faster Throughput in Digitization of Special Collections. Dublin, OH: OCLC, April 2011. http://www.oclc.org/research/publications/library/2011/2011-04.pdf.

[26] Mudge, Mark, Tom Malzbender, Carla Schroer, and Marlin Lum. “New Reflection Transformation Imaging Methods for Rock Art and Multiple-Viewpoint Display.” Proceedings from the 7th VAST International Symposium on Virtual Reality, Archaeology and Cultural Heritage (2006).

[27] Mudge, M., Schroer, C., Earle, G., Martinez, K., Pagi, H., Toler-Franklin, C., Rusinkiewicz, S., et al. “Principles and Practices of Robust, Photography-based Digital Imaging Techniques for Museums.” Proceedings from the 11th VAST International Symposium on Virtual reality, Archaeology and Cultural Heritage (2010).

Author Biography

Michael J. Bennett is Digital Projects Librarian & Institutional Repository Coordinator at the University of Connecticut. There he manages digital reformatting operations while overseeing the University’s institutional repository. Previously he has served as project manager of Digital Treasures, a digital repository of the cultural history of Central and Western Massachusetts and as executive committee member for Massachusetts’ Digital Commonwealth portal. He holds a BA from Connecticut College and an MLIS from the University of Rhode Island.