6-4-2010

Raw as Archival Still Image Format: A Consideration

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

F. Barry Wheeler
United States Library of Congress, bwhe@loc.gov

Raw as Archival Still Image Format: A Consideration

Michael J. Bennett; University of Connecticut Libraries; Storrs, CT/USA
F. Barry Wheeler; Library of Congress; Washington, DC/USA

Abstract

Source materials like fine art, over-sized, fragile maps, and delicate artifacts have traditionally been digitally converted through the use of controlled lighting and high resolution scanners and camera backs. In addition the capture of items such as general and special collections bound monographs has recently grown both through consortial efforts like the Internet Archive's Open Content Alliance and locally at the individual institution level. These projects, in turn, have introduced increasingly higher resolution consumer-grade digital single lens reflex cameras or "DSLRs" as a significant part of the general cultural heritage digital conversion workflow. Central to the authors' discussion is the fact that both camera backs and DSLRs commonly share the ability to capture native raw file formats. Because these formats include such advantages as access to an image's raw mosaic sensor data within their architecture, many institutions choose raw for initial capture due to its high bit-level and unprocessed nature.

However to date these same raw formats, so important to many at the point of capture, have yet to be considered "archival" within most published still imaging standards, if they are considered at all. Throughout many workflows raw files are deleted and thrown away after more traditionally "archival" uncompressed TIFF or JPEG 2000 files have been derived downstream from their raw source formats [1][2]. As a result, the authors examine the nature of raw anew and consider the basic questions, Should raw files be retained? What might their role be? Might they in fact form a new archival format space?

Included in the discussion is a survey of assorted raw file types and their attributes. Also addressed are various sustainability issues as they pertain to archival formats with a special emphasis on both raw's positive and negative characteristics as they apply to archival practices. Current common archival workflows versus possible raw-based ones are investigated as well. These comparisons are noted in the context of each approach's differing levels of usable captured image data, various preservation virtues, and the divergent ideas of strictly fixed renditions versus the potential for improved renditions over time. Special attention is given to the DNG raw format through a detailed inspection of a number of its various structural components and the roles that they play in the format's latest specification. Finally an evaluation is drawn of both proprietary raw formats in general and DNG in particular as possible alternative archival formats for still imaging.

Raw File Types & Attributes

According to Fraser and Schewe, "Fundamentally, a digital raw file is a record of the raw sensor data from the camera, accompanied by some camera-generated metadata... it's important to realize that 'digital camera raw' isn't a single file format. Rather it's a catchall term..." [3] Krogh further explains that raw image files contain, "1) The source image data - a "dump" of the information that the sensor gathered when the picture was taken. It's generally mosaiced data, and it has no inherent color balance. 2) An embedded preview - a JPEG conversion of the image so that you can see what it looks like. 3) Information about the photo - EXIF, private maker notes, possibly some subset of the IPTC metadata." [4] Broken down further, what exists in mosaiced raw image sensor data is simply a record of luminance values at each sensor element or what in essence is a grayscale image. Additionally, however, what is also recorded in the file are the characteristics of the camera manufacturer's color filter array or mosaic (usually arranged in a Bayer pattern) that is applied over the individual sensor elements. Thus the data is a representation of the scene colorimetery in the sensor's color space defined by its Bayer filter values, its individual photosite spectral sensitivities, and the camera system's processing as the data is drawn from the chip and buffered as a file. Final image production is then accomplished either through the manufacturer's own raw conversion software or through applications such as Adobe Camera Raw, Adobe Lightroom, or Bibble Lab's BibblePro. In each case, the converter's purpose is to combine both the luminance and color filter data to form a color image. This procedure is commonly known as demosaicing. Fraser illustrates the conversion process in the following [5][6]:

Figure 1. *An area array—each photosensor contributes one pixel to the image.*

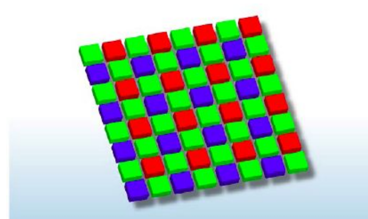

Figure 2. *In a Bayer pattern color filter array, each photosensor is filtered so that it captures only a single color of light: red, green, or blue. Twice as many green filters are used as red or blue because our eyes are most sensitive to green light.*

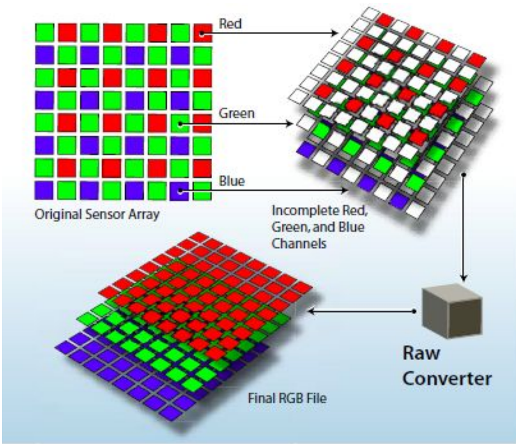

***Figure 3**. The raw capture is demosaiced and interpreted by a raw converter, using portions of the metadata embedded into the file at the time of capture, as well as algorithms in the conversion software. © 2004 Adobe Systems Incorporated. All rights reserved.*

In addition to the different filters that manufacturers employ, proprietary camera raw file types abound as well [7]. Among these, the storage of raw image data can vary among uncompressed, losslessly compressed, and lossy compressed options. Examples like Nikon's .nef offer either uncompressed or losslessly compressed choices depending upon camera model. Canon's .cr2 is exclusively losslessly compressed (original JPEG DPCM-based lossless mode) while their .sraw format employs lossy compression only [8][9][10][11].

Here, Krogh [12] provides a closer look at the final raw file after conversion:

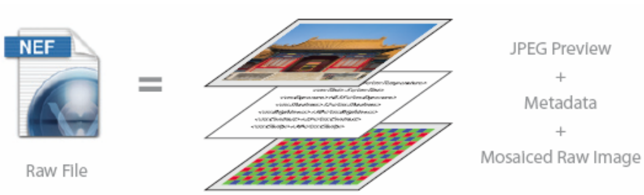

***Figure 4.** © 2007 Adobe Systems Incorporated. All rights reserved.*

Of particular note is that the metadata that is associated with the file can also contain rendering instructions. These instructions allow for what is known as *parametric image editing*. Here changes over settings like white balance, color temperature, exposure, cropping, etc. can be edited non-destructively through metadata directives rather than through traditional destructive image manipulations of *raster image editing* [13] where similar adjustments are irreversibly baked into the pixel data of rendered file types like TIFF. Parametric edits, on the other hand, do not change the original image data, they just re-interpret the source image data in a completely reversible manner. As a result raw files can contain not only a variety of embedded parametric renditions but any of these can either be exported into separate rendered file formats such as TIFF or reversed back to their default, latent state at the time of first raw file conversion. Finally, parametric edits are normally saved as XMP metadata that can live embedded in the raw file itself (e.g. DNG format), in a sidecar file (e.g. proprietary raw formats), or in a database [14].

DNG in Detail

In an attempt to standardize the current universe of multiple raw formats, Adobe created, published, and currently continues to update their "Digital Negative" DNG specification. To date, Adobe has also provided a free, universal license to their raw to DNG converter for all raw formats known to Adobe in addition to a complete and open DNG SDK for developers' use [15].

In creating a DNG file, the DNG converter reads a given proprietary raw file. The converter first “linearizes” the sensor data, then maps the sensor color space to the CIE XYZ color space with a D50 white point. The CIE XYZ color space is a standardized, device-independent space large enough to retain the entire range of colors captured by the sensor. This prepares the color data so that standard transforms can be used when rendering the file for a specific output device such as a monitor or printer [16][17]. At this point the linearized and standardized color data is then stored along with metadata that may describe the camera and capture system employed. Finally DNG can also accommodate a compressed copy of the entire original raw file embedded within the DNG architecture itself.

DNG is an extension of the TIFF 6.0 format and is compatible with the TIFF-EP subset [18]. In turn, DNG employs not only its own unique tag set but also currently uses five TIFF-EP tags. Included among these are *NewSubFileType* which allows DNG to store multiple image renderings as embedded preview images within the single DNG file. Such preview images can be uncompressed or compressed (either lossy or losslessly) as determined by the use of a *Compression* TIFF-EP tag value for a given preview. This flexibility allows for the future possibilities of direct preview image output for such roles as web delivery or high quality printing. Though not yet utilized in this fashion by current imaging software, such preview output could in certain instances replace the current need to repeatedly reinterpret the raw image data from scratch for such exporting tasks. In addition to its control over preview images, the repeatable *Compression* tag also determines if the original raw mosaic image data is stored uncompressed or losslessly compressed (lossless JPEG "old style," DPCM coding) within the DNG. Next, the *BitsPerSample* tag accommodates values from 8 to 32 bits per channel (bit depths of 24 to 96). Thus, DNG is able to handle both common DSLR camera models like Canon's EOS 5D which can capture up to 16 bits per channel (48-bit depth) on up to high-end camera backs. Lastly among the TIFF_EP tags, the presence of the *Orientation* tag is a DNG requirement and allows file browsers to perform lossless image rotation.

DNG is unique among raw formats in that it can contain accurate JPEG previews of parametric edits as they are made to a file. Proprietary raw types can currently only hold the original preview created by a given capture device [19][20]. Additionally among the DNG tags themselves, the newer (as of DNG v.1.2.0.0) *PreviewColorSpace* allows previews to be stored in color spaces beyond the older specification's single sRGB option. These alternatives include Gray Gamma 2.2 (grayscale), Adobe RGB, and ProPhoto RGB. The *PreviewApplicationName*, *PreviewApplicationVersion*, and *PreviewDateTime* tags all concern themselves with the possibility that over time users may employ numerous programs (including non-Adobe ones) to interact with a given DNG file and re-render its previews. *PreviewSettingsName* is currently used by Adobe Lightroom for its "Snapshots" feature where particular renderings and their previews can be named. Such an allowance for built-in image edit tracking grants users the ability to coherently label and revisit older renderings, to swap multiple renderings, or to better develop new ones all within the architecture of the single DNG file.

The *DNGPrivateData* and *MakerNoteSafety* tags provide storage of camera manufacturer's private data typically for use with their own raw converter software. Such information is unnecessary for other raw converters to work with the file. Data that may be held within these tags include proprietary "picture styles" codes that may be used by the camera manufacturer's software.

The *RawImageDigest* and *OriginalRawFileDigest* tags provide for the writing of a 16-byte, MD-5 hash for both the raw mosaic image data from the camera sensor and the original proprietary raw file respectively (if the user chooses to embed the proprietary file within the DNG file itself during conversion which the format supports). Compelling attributes of the *RawImageDigest* hash for digital preservation include the fact that it refers to just the original raw image data of the DNG file and not the previews or metadata areas. As a result, the hash, created automatically upon a DNG's initial creation, concerns itself only with the aspect of the file that should remain unchanged (raw image data) while ignoring those aspects of the file that are indeed designed to coherently change if the need arises (previews, metadata). In other words, any added parametric editing instructions or descriptive information updates written to the file's metadata will not invalidate the hash and falsely signal raw image data integrity problems. Additionally, because the hash is embedded within the file itself in the *RawImageDigest* tag, it can travel with the file and serve as a self-validation mechanism without the need of an external hash database. DNG is unique among present day file formats in its intelligent use of such an embedded hash in this manner. Currently Adobe's DNG Converter and Camera Raw engines both use the hash to verify files, but since the format is open, the tag may be used similarly by other software in future [21][22].

Support among camera manufacturers for DNG capture has steadily grown over time since the format's release. Today, models include Sinar, Hasselblad, Leica, and various Pentax bodies [23][24][25][26].

With the latest DNG specification has also arrived the concept of Opcode List tags which include, *OpcodeList1* through *OpcodeList3*. Through the use of these tags, specific processing steps like lens corrections, which are best moved off of the camera hardware's processing load, can instead be passed through more powerful computer workstation devices where DNG reader software normally runs. Additionally, such processes are advisedly performed after the image data has been demosaiced, a routine optimally accomplished in post-editing in order to maintain the full advantages of raw capture [27].

Finally, DNG files may be viewed in a variety of Adobe and non-Adobe image editing software [28][29]. Additionally, a DNG Codec has been developed by Adobe as a release candidate to allow Windows Vista users to view DNG files in the Windows Explorer and Photo Gallery [30][31].

Sustainability Issues

Here, issues of rendition must also be acknowledged within the context of raw. Given, for instance, the much followed workflow in current conversion labs, *archival master* > *processed master* > *derivatives*, current best practice within the cultural heritage community points to the importance of a fixed archival rendition of the original scene at the moment of capture [32][33][34]. Such rendering intents, accomplished under calibrated conditions by trained copy photographers, have been traditionally recorded in the fixed, de-mosaiced formats of uncompressed TIFF and JPEG 2000. Specifically, the conversion of the sensor data to TIFF is most often done by the scanner or camera processor. There, a demosaic algorithm selected by the manufacturer is first used to assign color values to each pixel, then sharpening is applied to counter the softening created by the device's anti-alias filter. Gamma and color transformations are then employed in order to put the image in a specific color space—usually the small gamut sRGB or the slightly larger gamut AdobeRGB. However, neither sRGB nor Adobe RGB can contain all of the colors or dynamic range of the original sensor data. Yet these transformations are irreversible in the creation of a rendered TIFF file. In contrast, a raw image file stores the capture device's original mosaic image data and can accommodate flexible parametric post-editing. This allows for the broader possibilities of improved image quality presently and into the future. With the growth of parametric image editing software there exist within raw the capacities for archiving not only a single rendering intent which may be limited by current technology but also future renderings of perhaps even greater fidelity as such technology and software progress [35][36].

Given the desire to possibly preserve a raw "negative" the following questions must be asked, *Is this image data sustainable? Can the image data and production of desired renditions over an extended period of time be maintained?* The Library of Congress provides a list of factors for assessing a given digital format for sustainability, along with an evaluation of many common formats already in use.

These factors include:

- "Disclosure — Disclosure refers to the degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content. Preservation of content in a given digital format over the long term is not feasible without an understanding of how the information is represented (encoded) as bits and bytes in digital files. ...however, what is most significant for this sustainability factor is not approval by a recognized standards body, but the existence of complete documentation, preferably subject to external expert evaluation. ..."
- "Adoption — Adoption refers to the degree to which the format is already used by the primary creators, disseminators, or users of information resources. ..."
- "Transparency — Transparency refers to the degree to which the digital representation is open to direct analysis with basic tools, including human readability using a text-only editor. ...Encryption is incompatible with transparency; compression

inhibits transparency. However, for practical reasons, some digital audio, images, and video may never be stored in an uncompressed form, even when created. Archival repositories must certainly accept content compressed using publicly disclosed and widely adopted algorithms that are either lossless or have a degree of lossy compression that is acceptable to the creator, publisher, or primary user as a master version. ..."

- "Self-documentation — Digital objects that are self documenting are likely to be easier to sustain over the long term and less vulnerable to catastrophe than data objects that are stored separately from all the metadata needed to render the data as usable information or understand its context. A digital object that contains basic descriptive metadata and incorporates technical and administrative metadata relating to its creation and early stages of its life cycle will be easier to manage and monitor for integrity and usability and to transfer reliably from one archival system to its successor system. ...Digital formats in which such metadata can be embedded in a transparent form without affecting the content are likely to be superior for preservation purposes. ..."
- "External Dependencies — External dependencies refers to the degree to which a particular format depends on particular hardware, operating system, or software for rendering or use and the predicted complexity of dealing with those dependencies in future technical environments. ..."
- "Impact of Patents — Patents related to a digital format may inhibit the ability of archival institutions to sustain content in that format. [...this does not mean the absence of all patent...] The core components of emerging ISO formats such as JPEG2000 and MPEG4 are associated with "pools" that offer licensing, preferably cost free, on behalf of a number of patent-holders. ..."
- "Technical Protection Mechanisms — To preserve digital content and provide service to users and designated communities decades hence, custodians must be able to replicate the content on new media, migrate and normalize it in the face of changing technology, and disseminate it to users at a resolution consistent with network bandwidth constraints. Content for which a trusted repository takes long-term responsibility must not be protected by technical mechanisms such as encryption, implemented in ways that prevent custodians from taking appropriate steps to preserve the digital content and make it accessible to future generations. ..." [37]

A cursory inspection of these requirements reveals that native camera raw formats do not meet the requirements for sustainability. Proprietary in nature, camera raw files are indeed not only tied to the specifics of the camera sensor size and arrangement, but also to the filters used to create color images, the mechanisms of light capture, and to the formatting of the data delivered to the memory storage system. Camera manufacturers do not completely disclose their imaging system specifications. Their systems are not transparent, self-documenting, nor are they free of technical protection mechanisms. In addition, such systems are completely dependent on the hardware and software of the camera manufacturer and may be protected by numerous patents. The Library notes, "The proprietary nature of raw formats, however, means that there is a risk that any given format will not be supported for the long term, especially if the manufacturer goes out of business." [38]

On the other hand, The Library evaluation of DNG is much more positive. Adobe published DNG version 1.0.0.0 in August of 2004, but this version of the specification did not contain the complete data for a proprietary raw file [39]. Adobe quickly corrected this flaw and published DNG Specification 1.1.0.0 in February, 2005 [40]. In turn the Library of Congress uses Version 1.1 as the foundation for their published sustainability evaluation of DNG [41]. DNG receives good marks on each of the factors. Thus, in the summary of all raw file formats, the Library states its preference as, "None at this writing, although normalization to DNG 1.1 may emerge as a preferred practice."

Two further issues might be considered here. First, DNG is currently not an ISO standard. As a result some in the community fear Adobe's current control [42]. Secondly, in an attempt to make DNG acceptable to the large camera manufacturers such as Canon, Nikon, and Sony, Adobe accepts proprietary data—possibly encrypted—in the DNG metadata, specifically in private tags, private IFDs, and/or a private MakerNote. The format does recommend, however, that manufacturers use the DNGPrivateData and MakerNoteSafety tags to better ensure the preservation of such proprietary data [43]. Large manufacturers continue to attempt to use features of their raw formats to commercial advantage. In turn, they may fear that DNG's broad adoption would promote a level playing field to the benefit of other manufacturers. Thus some camera specific, or special processing specific metadata may be hidden within the DNG file and unavailable to all conversion software which would make DNG not completely documented, not completely transparent, and perhaps not completely free of external dependencies and patent restrictions in the minds of some.

The issue of ISO standardization may be answered in two ways. As the Library of Congress' sustainability factors point out, the existence of complete documentation is more relevant to digital format sustainability than acceptance as an ISO standard. Adobe has not only published extensive documentation and an explicit patent license, they have also provided the free SDK and file converter to ensure that the DNG specification is as open and available as any ISO standard [44]. Adobe has also formally submitted the DNG specification to the ISO, and the .dng digital format may be incorporated into the next ISO Standard revision of TIFF/EP [45][46][47].

The second issue of camera manufacturers' adoption is less clear. As an example of the potential complexities involved, Nikon at one point in time encrypted the white balance metadata from their DSLR raw files and refused to provide to Adobe the algorithms required to unencrypt the data [48]. In this case, while white balance was still adjustable parametrically, the "as shot" white balance, normally an important element in providing the initial image rendition, was encrypted. As a compromise, Nikon now provides Adobe with a "mini-SDK" that reads the encrypted metadata and feeds it to the converter. This Nikon mini-SDK has been incorporated into Adobe's DNG SDK, so the encrypted data remains, but all conversion software can currently make use of the "as shot" white-balance metadata [49][50]. Similarly, additional issues may well arise in future raw file developments. Still, regardless of the hidden data, the raw to DNG converter can provide sufficient data to an image editing program so that an initial rendition can be created and the resulting image can be edited using widely available tools without requiring access to the hidden, proprietary metadata.

Current Archival Practices

There are two main workflows in current archival practice for the digitization of cultural heritage materials. In the traditional approach, project materials and objectives are evaluated to determine digitization specifications, primarily stated as spatial resolution and bit-depth metrics. A scanner (or camera) sensor captures light variations from the object. Then TIFF files are created in the scanner using the computer processing power and algorithms contained within the scanner to convert the sensor data into a bitmap representation or "rendition" of the object. This fixed rendition may be stored immediately or it may be post-processed to specific standards using more powerful computers and algorithms that can perform tonal adjustments, noise removal, sharpening and other processing to ensure the image capture meets project specifications. In either case, the TIFF image is a fixed rendition that is saved as a master file in some form of long-term storage. In this workflow, any in-scanner/camera processing and/or any post-processing permanently changes the pixels of the original capture. Once the image is fixed in its TIFF rendition, the content is frozen and no additional information can be extracted from the image capture; further processing merely changes specific pixel values [51][52].

At issue is the quality of the scanner data to a TIFF formatted image file conversion. Current National Archives [53], Library of Congress [54], and National Library of the Netherlands [55] guidelines suggest that resolution, tonal, noise, and color values be measured on controlled patches on targets designed for scanner image analysis. But recent unpublished testing at the Library of Congress by one of the authors (FBW) suggests that many images may have significant differences from the original source document. Analyzing images of standardized targets visually and using software-based numerical analysis shows a number of potential problems:

- Scans of grayscale step targets into rendered formats such as TIFF show many images have incorrectly placed white and black points. Scanned targets show merged tonal steps in both the darkest and lightest patches which result in significant loss of detail in both the dark and highlight tones. Tonal inaccuracies in the mid-scale patches also appear frequently in the target scans.
- Software computations following current ISO standards show that measured ISO resolution does not reach manufacturers' published resolution figures (which are frequently based on “sampling frequency” rather than on ISO standards). The analysis also indicates that some scanners and camera-backs over sharpen the rendered images they produce either through firmware routines built into the equipment or in external software processing of the raw sensor data. Sometimes operators can control this rendering process of the device, but sometimes it is an automated part of demosaicing the data from the Bayer array.
- Analysis also shows that some current scanners add measureable noise to rendered images.
- Additionally, for color images, scanners show inaccuracies in color reproduction and may clip or move significant color values. Most scanners internally convert the scene colors captured from sensor space directly to the limited gamut sRGB color space during the rendering processing that produces a TIFF file. The sRGB gamut is relatively small and some colors of the original may lie outside the gamut boundaries. When this happens, the outside colors may be moved into the sRGB color gamut and those colors already inside may be adjusted correspondingly. The problem is compounded because many scanners do not created and embed an ICC color profile making accurate representation of colors on monitors or printers difficult.

It should be noted that this situation is slowly improving. Newer scanners, camera backs, and the various software packages that can now drive such devices may allow the operator to choose a larger gamut color space, such as Adobe RGB (1998), and frequently do embed ICC profiles in the captured rendered images. But the problem remains when significant colors of the source document are outside the sRGB or AdobeRGB color spaces. Those colors are mapped into the designated color space—and colors inside the space adjusted according to the rendering intent. Such color changes are irreversible. In sum, immediate processing to rendered formats such as TIFF introduces a number of problems that can result in lower scanned image quality.

Another alternative might be to simply maintain the original source data if the camera or scanner will output a raw file. However, according to Steinmueller "Leaving your master image file in the camera RGB or scanner RGB color space [device color space] is usually a bad idea because those spaces are not gray balanced and not equal gamma for R,G,B. As you edit the master image file and raise or lower all RGB values the same percentage, you may unintentionally introduce a color cast or color crossover because the camera or scanner represents neutral colors as non-equal amounts of R,G,B. Much better to first use the camera profile or scanner profile to convert into a well-behaved editing space such as Lab [CIE L*a*b*] or ProPhoto RGB." [56]

The second workflow is currently employed primarily for large-scale digitization where hundreds of thousands—even millions—of books and documents are digitized. The largest and best known mass digitization project is the Google Book Scanning Project (performed at proprietary Google-owned imaging facilities). There, sensor image capture is immediately converted to JPEG 2000 format using lossy compression. This JP2 image file becomes the "archival master" which is saved in long-term storage. Lossy-compressed JPEG 2000 images save significant storage space at the minimal cost of small amounts of image detail. In this case, the project goal is to capture the documents' intellectual content, and the residual loss of certain artifactual value is acceptable in order to speed processing and save storage space [57][58][59][60].

Raw vs. Traditional “Positive” Archival File Types

When evaluating current raw-based workflows against more traditional archival strategies it is prudent to step back and consider the film negative and a print made from that negative.

The print is a *rendition* of the negative – but only one of an almost infinite number of different renditions that might be created from the negative. In many cases the print cannot present all the information in the negative. If the negative is examined on a light table with a loupe, it will be obvious that there is detail in the shadows or in the highlights that is not found in the print. And, if in ten years' time another print maker creates a new print from the same negative using a different enlarger, developer, and paper, and applies the printmaker’s craft, then the rendition will almost certainly be different.

The situation is similar with digital files. A DNG file, for example, contains the data taken directly from the camera sensor, plus embedded metadata describing the characteristics of the specific camera used for the capture. It may also contain metadata describing the processing instructions necessary to produce a rendition of the image. Such processing instructions can be used to produce a TIFF rendition file from the DNG just as one might follow the burning, dodging, developing, and toning darkroom notes to produce a traditional print from a negative. But producing a new print from an old negative using more modern chemicals, paper, enlarging lens and filters will produce a visibly different rendition even when the same darkroom notes are followed. Similarly, producing a TIFF "print" from a DNG file using a new or different raw converter will produce a visibly different rendition even though the same embedded processing instructions are used. Why does this happen?

The DNG demosaicing process requires several algorithms to produce an RGB image from sensor data, camera characteristics and capture parameters. Working from a DNG file using a powerful workstation and specially designed conversion software allows the use of sophisticated edge detection and retention algorithms that retain image detail lost during in-camera processing. When color is finally assigned to specific pixels, a nearest-neighbor algorithm can be augmented by more advanced techniques giving more accurate color values to each pixel. More accurate and less destructive parametric processing can be performed to remove noise from the original data rather than having such editing going on in a destructive manner on derived pixels that have already had noise introduced into their color values. Similarly, a variety of techniques can be used to retain highlight detail, and color values need not be moved or clipped to fit into the limited color gamut of sRGB.

Thus external workstation processing of DNG files may result in a significant improvement in imaging accuracy. Additionally, as technology improves over time, older DNG files can be reprocessed using updated algorithms and increased computational power unlike a static TIFF rendition where colors that have been moved or clipped cannot be restored, detail that has been lost cannot be recovered, and added noise cannot be separated from the fixed image values. In fact, through the use of new specific Opcode instructions, recent DNG converters have added the ability to apply corrections for known defects in specific lenses – new DNG to TIFF conversions can now correct both lens specific geometric aberrations and chroma aberrations. In turn, it is not unreasonable to expect that additional content and improved renditions from the original raw mosaic sensor data will become available from older DNG files as conversion software improves.

But should not there be the ability to revert to an original rendition? Again, a consideration of film negatives is helpful. While the modern workflow will produce a visibly different rendition of an older negative, with effort the darkroom worker can locate older paper and developers, setup the enlarger to mimic an older enlarger, and with special effort produce a rendition that is visibly similar to the original. In the same way, most raw converters can be set to interpret the original image creation parameters to create a rendition similar to the original. Thus, just as both prints and negatives in analog formats are archived, consideration should be given to archiving original raw mosaic sensor data (preferably in DNG) as well as the more traditional TIFF or JPEG rendition of a given image.

Summary

Currently, the TIFF image format is considered the most appropriate for long-term archival storage. However, as the authors have noted the immediate conversion to TIFF may create flaws and inaccuracies of tonal values and color. Because this conversion is irreversible all flaws are “baked” permanently into the file.

Raw formats are an increasingly intriguing option for digital preservation of still images. As a possible archival format space they offer many advantages. These include access to the raw mosaic sensor data, more control over the original rendering process, greater bit-depth, wider color gamut options, broader usable dynamic range, non-destructive parametric editing, and storage requirements of roughly one third the size of such uncompressed rendered specifications as TIFF. However, proprietary raw formats have key drawbacks. One is that they are generally poor at storing additional custom metadata, an attribute of increasing importance in digital asset management [61][62][63]. Secondly, but perhaps of most importance, is that proprietary raw formats do not meet a number of basic criteria for sustainability. These include both transparency and self-documentation [64]. As a result they are acutely susceptible to being unreadable and orphaned in the future.

However, the open and fully documented DNG raw standard retains the common virtues of raw formats while also offering additional archival value. Unlike proprietary raw, DNG maps sensor-specific color space into the standardized CIE XYZ space. This allows users the flexibility to make a variety of subsequent working color space choices based upon a known device-independent standard. DNG is fully XMP-compliant and can use an embedded XMP space for both descriptive and technical metadata in a flexible and ultimately extensible way. Unlike proprietary raw with its separate sidecar file, DNG can also write parametric edits to embedded XMP which allows such rendering instructions to not only be completely portable along with the file itself but also allows such instructions to be more easily managed. Accurate embedded JPEG previews based upon parametric editing adjustments are also unique to the format. Additionally, DNG has the ability to embed custom camera profiles that allow for refined compensation among cameras of the same make and model [65]. DNG furthermore utilizes an internal MD-5 hash for its original mosaic image data. Thus the format is uniquely able to self-validate, a critical asset for well-managed archival file preservation. In the case of DNGs that have been converted from proprietary raw files, the format offers the additional option to embed a zip-compressed copy of the complete original raw file along with a second embedded hash for this data as well. In fact this is currently the only way in which proprietary raw files can be validated beyond a separately created hash database [66][67][68][69].

Significantly, all raw formats—among which DNG is best poised for sustainability farther into the future—possess the potential to improve over time. Because they do not statically render and fix image information but instead act as a container for raw capture data and metadata, they have the ability to take advantage of ongoing advances in rendering software that can make fuller use of such source data in richer and more refined ways. As a result, image quality from older DNG files can continue to improve into the future through advances like more sophisticated shadow detail extraction algorithms, noise reduction, and lens correction capabilities. This act of re-tapping currently unused original sensor information is impossible with rendered formats as the veins of source data are forever severed when pixels are fixed and original raw files are deleted.

DNG is currently adopted as an archival master format at The National Gallery of Art, Washington, D.C. and at The Art Institute of Chicago [70]. The National Gallery in particular acknowledges the specification's high end image data storage, growing adoption, and single file workflow practicality (archival masters > production derivatives) [71]. Yet, they also report present difficulties as well. These include consistent configuration among differing DNG editing software and the tendencies of these programs, when DNG files are moved among collaborative users, to overwrite the original photographer's development adjustments with the default settings of the recipient if such default settings are not fastidiously checked. In an enterprise-level, multi-workstation processing environment this can add an additional layer of technical oversight [72]. Such complexity contrasts with the relative simplicity of TIFF raster editing and handling. As a result DNG training and workflow implementation may be somewhat more difficult for new adopters even in light of the format's rapidly improving software support and smoother functionality. As a baseline, familiarity with the concepts of parametric editing is necessary to confidently sustain DNG files while the format and its tools continue to mature.

References

[1] Stanley Smith et al., “Speed the Plow: Rapid Capture Digital Workflow Handout,” in (presented at the The Museum Computer Network Annual Conference, Portland, OR, 2009), (retrieved 1.20.10), http://www.mcn.edu/conference/mcn2009/Smith_Getty_Rapid_Capture_Project_Photographs.pdf.

[2] Matt Pearson, Carolyn Frisa, and Monique Fischer, “How we got "there" from here: an overview of the design and implementation of a digital documentation photography studio,” Journal of the American Institute for Conservation 48, no. 2 (Summer 2009).

[3] Bruce Fraser and Jeff Schewe, Real world Camera Raw with Adobe Photoshop CS3 (Berkeley, Calif. : Peachpit ; London : Pearson Education [distributor], 2008) pg. 2.

[4] Peter Krogh, The DAM book : digital asset management for photographers, 2nd ed. (Beijing ; Cambridge [Mass.] : O'Reilly, 2009) pg. 42.

[5] Bruce Fraser, “Understanding Digital Raw Capture,” (retrieved 08.17.09), http://www.adobe.com/products/photoshop/pdfs/understanding_digitalrawcapture.pdf.

[6] Bruce Fraser and Jeff Schewe, Real world Camera Raw with Adobe Photoshop CS3 (Berkeley, Calif. : Peachpit ; London : Pearson Education [distributor], 2008) pg. 3.

[7] “Raw image format - Wikipedia, the free encyclopedia,” (retrieved 10.11.09), http://en.wikipedia.org/w/index.php?title=Raw_image_format&oldid=319218981.

[8] “Canon EOS 5D Mark II Review: 12. Menus: Digital Photography Review,” (retrieved 08.09.09), http://www.dpreview.com/reviews/canoneos5dmarkii/page12.asp.

[9] “Rawzor - Supported Formats and Cameras,” (retrieved 08.13.09), http://www.rawzor.com/formats/.

[10] William Pennebaker and Joan Mitchell, Jpeg: Still Image Data Compression, 8th ed. (Van Nostrand Reinhold, 1992), pg. 78.

[11] “Inside the Canon RAW format version 2, understanding .CR2 file format and files produced by Canon EOS Digital Camera,” (retrieved 08.13.09), http://lclevy.free.fr/cr2/.

[12] Peter Krogh, Non-Destructive Imaging: An Evolution of Rendering Technology, (retrieved 10.14.09), http://www.adobe.com/digitalimag/pdfs/non_destructive_imaging.pdf.

[13] Patti Russotti and Richard Anderson, Digital photography best practices and workflow handbook: a guide to staying ahead of the workflow curve. (Burlington, MA: Focal Press, 2010) pg. 53.

[14] Peter Krogh, Non-Destructive Imaging: An Evolution of Rendering Technology, (retrieved 10.14.09), http://www.adobe.com/digitalimag/pdfs/non_destructive_imaging.pdf.

[15] “Adobe - Digital Negative (DNG),” (retrieved 08.05.09), http://www.adobe.com/products/dng/.

[16] Sabine Süsstrunk,, “Color Encodings for Image Databases,” Proceedings IS&T/SPIE's Electronic Imaging: Internet Imaging Vol. 4672 (2002) pg. 179-185.

[17] Sabine Süsstrunk,, Robert Buckley, and Steve Swen, “Standard RGB Color Spaces,” Proceedings from the IS&T/SID 7th Color Imaging Conference Vol. 7 (1999) pg. 127-134.

[18] Adobe Developers Association, “TIFF Revision 6.0 Final-June 3, 1992,” (retrieved 10.11.09), http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf.

[19] Patti Russotti and Richard Anderson, Digital photography best practices and workflow handbook: a guide to staying ahead of the workflow curve. (Burlington, MA: Focal Press, 2010) pg. 174-175.

[20] Peter Krogh, The DAM book : digital asset management for photographers, 2nd ed. (Beijing ; Cambridge [Mass.] : O'Reilly, 2009) pg. 50, 57.

[21] Ibid.pg. 52-60.

[22] “dng_1_3_0_0_spec_June_2009,” (retrieved 08.13.09), http://www.adobe.com/products/dng/pdfs/dng_spec_1_3_0_0.pdf.

[23] “Adobe - Adobe Photoshop CS4: Digital camera raw file support,” (retrieved 08.17.09), http://www.adobe.com/products/photoshop/cameraraw.html.

[24] “Sinar eXposure™ 6.0 Software Tips,” (retrieved 08.14.09), http://www.sinarcameras.com/newsletter/003_29042008/attachments/DYK.pdf.

[25] “Shutterbug: Sinar eXposure,” (retrieved 08.14.09), http://www.shutterbug.com/news/071408sinar/.

[26] “Luminous Landscape Forum > SINAR RAW file conversion to DNG,” (retrieved 08.14.09), http://www.luminous-landscape.com/forum/lofiversion/index.php/t27909.html.

[27] “dng_1_3_0_0_spec_June_2009,” (retrieved 08.13.09), http://www.adobe.com/products/dng/pdfs/dng_spec_1_3_0_0.pdf.

[28] Barry Pearson, “Products that support DNG in some way,” (retrieved 10.29.09), http://www.barrypearson.co.uk/articles/dng/products.htm#enduser.
[29] Barry Pearson, “Adobe products that support DNG,” (retrieved 10.29.09), http://www.barrypearson.co.uk/articles/dng/products_adobe.htm.
[30] “DNG Specification and Vista DNG Codec - Lightroom Journal,” (retrieved 10.24.09), http://blogs.adobe.com/lightroomjournal/2008/05/dng_specification_vista_dng_co.html.
[31] “DNG Codec - Adobe Labs,” (retrieved 08.13.09), http://labs.adobe.com/wiki/index.php/DNG_Codec.
[32] “Technical Information (American Memory from the Library of Congress),” (retrieved 11.04.09), http://memory.loc.gov/ammem/about/techIn.html.
[33] “Guidelines for Digitizing Archival Materials for Electronic Access,” (retrieved 10.07.09), http://www.archives.gov/preservation/technical/guidelines.html.
[34] Franziska Frey and James Reilly, “Digital Imaging for Photographic Collections Foundations for Technical Standards,” 2006, (retrieved 10.05.09), http://www.imagepermanenceinstitute.com/shtml_sub/digibook.pdf.
[35] Patti Russotti and Richard Anderson, Digital photography best practices and workflow handbook: a guide to staying ahead of the workflow curve. (Burlington, MA: Focal Press, 2010) pg. 72.
[36] Jeff Schewe, “PhotoshopNews: Photoshop News and Information » Archive » Digital Preservation,” (retrieved 08.13.09), http://photoshopnews.com/2005/05/11/digital-preservation/.
[37] “Sustainability Factors,” (retrieved 09.14.09), http://www.digitalpreservation.gov/formats/sustain/sustain.shtml.
[38] “Camera Raw Formats (Group Description),” (retrieved 10.25.09), http://www.digitalpreservation.gov/formats/fdd/fdd000241.shtml.
[39] “dng_1_0_0_0_spec_Sept_2004,” (retrieved 10.08.09), http://web.archive.org/web/20041204034106/www.adobe.com/products/dng/pdfs/dng_spec.pdf.
[40] “dng_1_1_0_0_spec_Feb_2005,” (retrieved 10.08.09), http://web.archive.org/web/20050724102600/www.adobe.com/products/dng/pdfs/dng_spec.pdf.
[41] “Adobe Digital Negative (DNG), Version 1.1,” (retrieved 10.05.09), http://www.digitalpreservation.gov/formats/fdd/fdd000188.shtml.
[42] “Two Paths Leading Nowhere | LibRaw,” (retrieved 09.14.09), http://www.libraw.org/articles/2-ways-to-nowhere.html.
[43] “dng_1_3_0_0_spec_June_2009,” (retrieved 08.13.09), http://www.adobe.com/products/dng/pdfs/dng_spec_1_3_0_0.pdf.
[44] “Adobe - Digital Negative (DNG).” (retrieved 10.05.09), http://www.adobe.com/products/dng/.
[45] “Tag Image File Format / Electronic Photography - Wikipedia, the free encyclopedia,” (retrieved 10.12.09), http://en.wikipedia.org/wiki/Tag_Image_File_Format_/_Electronic_Photography.
[46] “John Nack on Adobe: DNG submitted to the ISO,” (retrieved 08.13.09), http://blogs.adobe.com/jnack/2008/05/dng_submitted_t.html.
[47] “Adobe seeks International recognition for DNG: Digital Photography Review,” (retrieved 08.13.09), http://www.dpreview.com/news/0805/08051503adobeDNG.asp.
[48] Photoshop News Editorial Staff, “PhotoshopNews: Photoshop News and Information » Archive » Nikon encrypts D2X white balance metadata,” PhotoshopNews.com, April 17, 2005, (retrieved 10.19.09), http://photoshopnews.com/2005/04/17/nikon-d2x-white-balance-encryption/.
[49] “What to Make of Adobe's New DNG Raw Format - Photo Tips @ Earthbound Light,” Earthbound Light, September 26, 2005, (retrieved 10.19.09), http://www.earthboundlight.com/phototips/adobe-dng-raw.html.
[50] “NikonUSA Press Release on DNG,” September 2, 2005, (retrieved 10.19.09), http://www2.prnewswire.com/cgi-bin/micro_stories.pl?ACCT=130907&TICK=NIKON&STORY=/www/story/09-05-2005/0004100183&EDATE=Sep+5,+2005.
[51] “Technical Information (American Memory from the Library of Congress).” (retrieved 11.04.09), http://memory.loc.gov/ammem/about/techIn.html.
[52] “Guidelines for Digitizing Archival Materials for Electronic Access.” (retrieved 10.07.09), http://www.archives.gov/preservation/technical/guidelines.html.
[53] Ibid.
[54] “Technical Information (American Memory from the Library of Congress).” (retrieved 11.04.09), http://memory.loc.gov/ammem/about/techIn.html.
[55] “Metamorfoze |,” (retrieved 11.04.09), http://www.metamorfoze.nl/en/methodiek/conversion.html.
[56] Uwe Steinmueller, “Color Management for Photographers #006 @Digital Outback Photo,” (retrieved 02.01.10), http://www.outbackphoto.com/color_management/cm_06/essay.html.
[57] Robert Buckley, “JPEG2000 - A Practical Digital Preservation Standard?,” February 2008, (retrieved 10.07.09), http://www.dpconline.org/docs/reports/dpctw08-01.pdf.
[58] Oya Rieger, “Preservation in the Age of Large-Scale Digitization: A White Paper,” (retrieved 10.08.09), http://www.clir.org/pubs/reports/pub141/pub141.pdf.
[59] Jennifer Weintraub and Melissa Wisner, “MASS Digitization at Yale University Library: Exposing the Treasures in Our Stacks,” Computers in Libraries (October 2008), (retrieved 10.07.09), http://findarticles.com/p/articles/mi_hb6365/is_200810/ai_n32297344/?tag=content;col1.
[60] Stephen Chapman et al., “Page Image Compression for Mass Digitization,” (retrieved 10.07.09), http://preserve.harvard.edu/massdig/hul_study/IST_PageImageCompression_preprint.pdf.
[61] Patti Russotti and Richard Anderson, Digital photography best practices and workflow handbook: a guide to staying ahead of the workflow curve. (Burlington, MA: Focal Press, 2010) pg. 32, 46, 50,102
[62] Richard Anderson, “Raw vs. Rendered | dpBestflow,” dpBestflow, (retrieved 11.17.09), http://dpbestflow.org/node/264.
[63] Peter Krogh, The DAM book : digital asset management for photographers, 2nd ed. (Beijing ; Cambridge [Mass.] : O'Reilly, 2009) pg.49-52, 111.
[64] “Sustainability Factors,” (retrieved 09.14.09), http://www.digitalpreservation.gov/formats/sustain/sustain.shtml.
[65] Patti Russotti and Richard Anderson, Digital photography best practices and workflow handbook: a guide to staying ahead of the workflow curve. (Burlington, MA: Focal Press, 2010) pg.152, 170.
[66] “dng_1_3_0_0_spec_June_2009,” (retrieved 08.13.09), http://www.adobe.com/products/dng/pdfs/dng_spec_1_3_0_0.pdf.
[67] Patti Russotti and Richard Anderson, Digital photography best practices and workflow handbook: a guide to staying ahead of the workflow curve. (Burlington, MA: Focal Press, 2010) pg. 84.
[68] Richard Anderson, “Archive File Formats | dpBestflow,” dpBestflow, (retrieved 11.17.09), http://dpbestflow.org/node/371.
[69] Peter Krogh, The DAM book : digital asset management for photographers, 2nd ed. (Beijing ; Cambridge [Mass.] : O'Reilly, 2009) pg. 56, 248, 250-252
[70] Stanley Smith et al., “Speed the Plow: Rapid Capture Digital Workflow Handout,” in (presented at the The Museum Computer Network Annual Conference, Portland, OR, 2009), (retrieved 1.20.10), http://www.mcn.edu/conference/mcn2009/Smith_Getty_Rapid_Capture_Project_Photographs.pdf.

[71] Peter Dueker, “Personal communication with the National Gallery's Peter Dueker, 11/06/09,” November 6, 2009.
[72] Peter Dueker, “Personal communication with the National Gallery's Peter Dueker, 11/16/09,” November 16, 2009.

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.

F. Barry Wheeler is Digital Projects Coordinator in the Office of Strategic Initiatives at the Library of Congress. Previously he has worked as a web services specialist at the FDA Medical Library, and as a Professor of Library Science at The Catholic University of America. He holds a Masters in Library Science from Catholic University and a Doctorate in Instructional Computing from The University of Tennessee.