7-1-2005

The use of technical metadata in still digital imaging by the newspaper industry

Howard Vogl

The Use of Technical Metadata in Still Digital Imaging by the Newspaper Industry

By
Howard Vogl

A thesis submitted in partial fulfillment of the
Requirements for the degree of Master of Science in the
School of Print Media in the College
of Imaging Arts and Sciences of the
Rochester Institute of Technology

July 2005

Primary Thesis Advisor: Professor Franziska Frey
Secondary Thesis Advisor: Professor Douglas Ford Rea

Acknowledgements

I would like to thank the faculty and staff at the Rochester Institute of Technology for their guidance and support throughout this project. Franziska Frey, primary thesis advisor; Douglas Ford Rea, secondary thesis advisor; Twyla Cummings, Graduate Coordinator; Professor Michael Kleper; Frank Cost, Associate Dean of the College of Imaging Arts and Sciences. Additionally, I would like to thank Adrienne McHargue for assistance with transcription, and Norma Fleck for handling financial arrangements.

Also, I would like to thank the newspaper professionals who freely gave their valuable time and knowledge to allow the completion of this research. Lori Craig, Steve Terrillion, and Miles Weissman from *USA Today*; Joan Bever, Kelly B. Hagen, and Rick Berry from the *Chicago Tribune*; Kevin Connor and Tim Fitzsimons from *The Washington Post*; Toby Usnik and Tara House from *The New York Times;* Dennis Floss from the Rochester *Democrat & Chronicle*.

Furthermore, I am grateful to the Scripps Howard Foundation for their generous financial assistance that allowed the completion of this project.

Finally, I would like to thank my wife, Rita, for providing the patience, understanding, and faith to make it all possible.

Table of Contents

List of Tables

List of Figures

Abstract

Newspapers are increasingly capturing images digitally. Included with these digital files is technical information about the conditions of the image and the conditions surrounding image capture. Technical metadata has the potential to be a valuable resource in image reproduction, management, and archiving. Nevertheless, even though digital devices capture a large amount of technical metadata, the use of such data in the digital imaging workflow is not widespread.

The use of technical metadata requires a uniform set of technical metadata standards and an open encoding scheme to embed data. From their inception, image file formats, such as TIFF and JPEG, have allowed the inclusion of technical metadata tags. The Exif schema has extended the metadata inclusion capabilities of both of these formats. Additionally, XML has emerged as a standard for users to add metadata to image files. Consequently, organizations such as the World Wide Web Consortium and Adobe Systems all support XML. Moreover, organizations such as the Digital Imaging Group (DIG35) and the National Information Standards Organization (NISO) are defining standards for technical metadata inclusion.

The purpose of this study was to answer two fundamental questions about technical metadata in the newspaper industry. First, it assessed the ability of technical metadata to improve the newspaper digital imaging workflow; and second, it determined how technical metadata could be used to preserve the integrity of newspaper digital

images.

This study examined five large newspaper organizations: *The Chicago Tribune, The New York Times, The Rochester Democrat & Chronicle*, *USA Today*, and *The Washington Post*. Based on interviews and questionnaire responses, each organization's use of technical metadata in the digital imaging workflow was examined through case studies. Interviews were conducted with those individuals responsible for image capture, adjustment, database management, and output. Furthermore, participants were asked to rank the importance of selected fields of technical metadata through a questionnaire.

It was found that the use of technical metadata classified by NISO as Basic Image Parameters, which includes file size, type, and compression, was universal in newspaper organizations. The use of Image Creation metadata was not widespread with the exception of two fields that established date and time of capture and assigned each image a unique identifier. Image Performance Assessment metadata, such as test targets, was not widely used except by *The Rochester Democrat & Chronicle*. Change History fell victim to the short cycle time in the newspaper industry, and for the most part, a history of change was kept at various handoffs in the digital workflow through versioning.

The use of technical metadata to improve the digital workflow, to an extent, was at cross-purposes with newspapers' need to visually examine each image to determine its usefulness. However, software designed to visually present technical metadata through a well designed graphic user interface was popular. It appeared that technical metadata had the potential to benefit newspapers when repurposing images for other media. Additionally, large newspaper organizations were creating their own image databases;

while the use of technical metadata in these archives was unclear, it would be prudent to include too much technical metadata, rather than too little.

The foremost concern of all organizations was preservation of the editorial integrity of the image. Newspapers defined editorial integrity as the ability to capture as much detail of an event as possible, and then present that information to their readers in a truthful, unambiguous way. Research pointed out that image reproduction quality was only one of a series of variables that determined newspaper image quality.

With the advent of digital photography, photographers are editing more in the field, and as a result they are making decisions regarding image content. The use of technical metadata has the potential to provide greater tractability of these outtakes.

Additionally, the industry is moving toward the Camera Raw file format to acquire image data that is unprocessed by camera software. The adjustment of Camera Raw files through a GUI, and their subsequent conversion to another file format, represented a de facto use of technical metadata to preserve editorial integrity.

Chapter 1

Description of the Problem

Statement of the Problem

Newspapers are increasingly capturing images digitally. Technical information about the conditions of the image and the conditions surrounding the image capture is often included with these digital files. This information, called technical metadata, has the potential to be a valuable resource in image reproduction, management, and archiving. Nevertheless, even though digital devices capture a large amount of technical metadata, its use in the digital imaging workflow is not widespread. To take full advantage of the quality and the efficiency of digital imaging, technical metadata needs to be incorporated into the digital imaging workflow.

It is evident that percentage of newspaper images will become valuable archives. Technical metadata could provide valuable historical information about the circumstances of image capture. Additionally, many newspapers are considering digitizing their existing photographic collections to preserve original documents and to make digital versions of these documents available to the public. The capture and the preservation of technical metadata is vital in allowing access and preserving the technical history of these documents.

Background and Significance

Analog and Digital Workflows

In traditional film photography, newspaper photographers chose equipment based on their needs for image quality and personal mobility. The choice of film was based on a limited number of options. Prints were made from negatives, and results were viewed without the aid of a rendering device. This system worked well in its time, but it was not without disadvantages. Analog photography was time consuming; therefore, limiting creativity in time-sensitive publications such as newspapers. Saldago (1998) comments that between travel and processing time, newspaper photographers using film spent less than half of their working day actually shooting photographs. In addition, the process used to make color separations from film was expensive. Chung (2001) observes, that in the past, color separations were made using a closed-loop system where the input and output devices were fixed using vendor-specific technology that required considerable operator skill.

One advantage of the film workflow was that it produced a tangible photographic print. Not only was this photographic print used in the newspaper reproduction process, but it also provided archivists with a physical record of the captured event. It was common practice to attach both descriptive and technical data to this print. However, unlike photographic prints, image files are not tangible, so information about image capture is not normally included with the digital file. This need has been met by the inclusion of metadata in image files. In the last decade, the capture of metadata by digital

recording devices has increased considerably; however, its use in image processing and archiving has not.

Digital imaging technology has been rapidly adopted because of its immediacy, its lower cost, and its creative options. However, to a great extent, the digital imaging workflow still mimics that of film. As Cost (2004) points out, the method in which new technologies are used changes slowly because their use is based on cultural habits. It is because of these cultural habits that the use of technical metadata in digital imaging has lagged behind its capture.

Interest in the Study

The researcher is interested in fully understanding the capture, management, and reproduction of digital images. The study of technical metadata assists in this purpose because it represents a common denominator in digital workflows. Therefore, this study intends to examine the capture and management of technical metadata, and how it can be used throughout the digital imaging workflow. The newspaper industry is an excellent medium for this study because newspapers capture, process, publish, and archive vast quantities of images quickly.

This study is significant because digital imaging blurs the boundary between image creation and its source. In theory, a digitally captured image is an image that is ready to reproduce. In reality, the direct reproduction of digital images rarely yields acceptable results because digital images need to be repurposed for their intended output. The acceptable output of a digital image involves considerations such as file size, format,

compression, color space, and image capture conditions. The use of technical metadata in digital imaging has the potential to reduce the boundary that exists between image creation and image use, while providing traceability to the original image capture.

Chapter 2

Literature Review

Introduction

The use of metadata in digital workflows promises to revolutionize image processing and archiving. This literature review begins with a description of image file metadata categories. Next, it examines the methods by which metadata is included in these files. The review then examines the development of standards for the inclusion of technical metadata in image files. Last, it explores the role of technical metadata in the newspaper industry.

Types of Metadata

Metadata is commonly defined as the data about the data. Being more descriptive, Baca (1998) from the Getty Information Institute notes that there are five types of metadata. **Administrative** metadata is used in the management of digital resources, such as the location and version of a digital archive. **Descriptive** metadata specifies a digital resource to speed retrieval. **Preservation** metadata describes the processes digital resources undergo, such as refreshing, migration, and integrity checks. **Use metadata** tracks the use of the digital resource. **Technical metadata** documents digitization information, such as capture conditions, capture device, image resolution, and image color quality (p. 3).

Associating Metadata with Image Files

There are several ways that technical metadata can be associated with an image file. Conceptually, there are two ways to associate technical metadata with an image file. The first method is embedding the data in the file; the second method is linking the data to the file. Image files including metadata consist of three parts. The first part of the file is the header, which contains essential information about the image file, as well as the location, or byte offset, of metadata tags within the file. The second part of the file is the actual image data, and the third part consists of descriptive and technical metadata. The linked file is similar in structure, except that metadata is externally attached to the file. Figure 1 shows a conceptual view of both types of image files (DIG 35, 2000, p. 5).

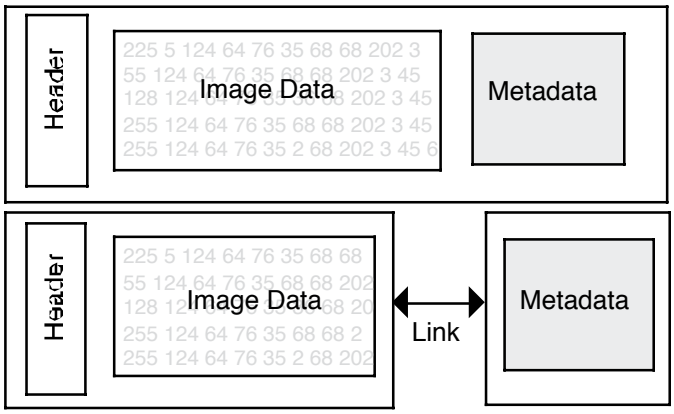


Figure 1. Association of metadata with image files.

Subdividing in a different way, McDowell acknowledges technical metadata that is a tag in an image file, and technical metadata that is an XML document attached to an image file. However, he adds a third form; technical metadata that exists as a separately recorded document (McDowell, D., personal communication, August 19, 2004).

Separately Recorded Technical Metadata

Separately recorded technical metadata is usually analog data that is associated with an image. Technical metadata as a separate document can be as simple as information written on the side of the negative that is relevant to producing a print. Photographers have used this form of technical metadata for many years. Kodak's Advantix photo system is an interesting example of analog technical metadata linked to an image. In the Advantix system, technical metadata is recorded on the margin of the film negative using magnetic technology in order to improve the reproduction of prints in consumer photography. As Kodak (2000) points out "Some Advantix cameras even allow the film to communicate with the photofinisher's equipment, resulting in automatic adjustments for the best prints."

Embedded Technical Metadata

TIFF. Another method of including technical metadata is through the use of tags in the image file. Adobe's Tagged Image File Format (TIFF) is the foremost example of using tags in image files. The TIFF file format has a long history, and as Cost (1997) notes TIFF was first created by Microsoft as a way to store pixel-based data (p. 36). In TIFF Revision 6.0 (1992) Adobe notes "The first version of the TIFF specification was published by Aldus Corporation in the fall of 1986, after a series of meetings with various scanner manufacturers and software developers" (p. 4).

TIFF has evolved as a popular image file format because of its image quality and cross-platform capabilities. From its inception, TIFF has been an image file format that

has allowed the inclusion of technical metadata. TIFF specification 6.0 standardized the TIFF file format and the inclusion of metadata with image files. Adobe (1992) claimed in TIFF Revision 6.0 that TIFF Final is an extensible image file format that is able to accept many forms of image data, is operating system independent, allows several compression schemes, and allows for the inclusion of metadata (pp. 4-5).

The TIFF file structure begins with a file header, containing information about the image, and describing the byte offset of the actual image. Additionally, the header describes the type and location of metadata tags. As early as the 1992 TIFF specification Adobe noted, "TIFF writers are allowed to write optional fields such as Make, Model, and Date Time" (p. 26). In addition to descriptive metadata, TIFF allows the inclusion of technical metadata, including bits per sample, compression, host computer, scanner manufacturer and model, image resolution, and software used to create the image (Adobe, TIFF, 1992, pp. 28-41).

JPEG2000/Part6. A newer file format containing tags is JPEG2000/Part6. JPEG2000/Part6 is an emerging file format that has been slated to replace the fifteen-year-old JPEG file format. Historically, the JPEG image file format was popular because of its high compression ratios and subsequently small file sizes. However, because of the nature of its compression scheme, JPEG has always been considered a lossy format. In contrast, JPEG2000/Part6 allows users to create a lossy file, apply different levels compression to different parts of a document, or create a lossless file. Jung & Zellmann (2004) describe JPEG200/Part6 as " a step in the direction of the next generation of

compression algorithms where scanned images are first separated into homogeneous parts and then individually compressed and transmitted together with mask information indicating how to combine the individual parts" (p. 281). JPEG2000/Part6 has the advantage of producing a smaller file than TIFF without the loss of image quality. Like TIFF, JPEG2000/Part6 also allows the inclusion of technical metadata regarding the image creation, image capture device, and capture device settings. Importantly, *Photoshop CS*, released in 2003, allows the user the option of including previously captured metadata when migrating to JPEG2000.

Camera Raw Files. An increasingly popular file format for capturing still digital images is the Camera Raw file format. When a digital camera captures a scene the camera's onboard software processes the image to an extent. Normally, camera software corrects for white balance and color, and then performs some sharpening before saving the digital file as a JPEG. Not only does the camera software make imaging decisions, but also the initial digital capture is saved in the lossy JPEG format. Capturing a digital image as a Camera Raw file not only circumvents onboard image processing, but also saves the image in the lossless Camera Raw file format (Long, 2003). Adobe (2003) describes the Raw format saying, "The image information is directly captured from the camera's CCD or CMOS without filters and adjustments applied by the camera; a digital negative" (p. 1). Understandably, the concept of a digital negative, that allows complete creative control, is appealing to newspaper photographers.

Camera Raw camera files are specific to the camera used for image capture. Therefore, to work with Camera Raw files the image adjustment software needs to be able to read Raw files from several cameras. Adobe provides Camera Raw plug-in software that allows the user to open and adjust Raw files derived from most popular cameras. In Adobe's plug-in, the user adjusts the Raw file and then saves it in the desired image format. Technical metadata about the scene and the camera settings serves as a reference for the photographer to make these adjustments. The adjustments made, including a log of changes, are recorded as technical metadata. The distinction between Camera Raw and JPEG file types is important; in JPEG, technical metadata describes conditions that have altered by camera software, while in Camera Raw, technical metadata reflects information about capture conditions and camera settings that have not been altered by camera software.

Linked Technical Metadata

XML and RDF. Metadata that is added or modified after image capture creates a unique challenge because of the variety of systems and applications used to create metadata. To encourage a uniform and open encoding scheme for metadata, the World Wide Web Consortium (W3C) supports the use of Extensible Markup Language [XML] and Resource Description Framework [RDF] (W3C, 2003). Because of its cross platform capability, XML has become the de facto standard for adding technical metadata to image files. Kodak notes, "Emerging standards, such as DIG35 and JPEG2000, are creating

metadata storage specifications based on XML" (Kodak, *Technology Highlights, Metadata*).

The beauty of XML is its independence. W3 Schools declares, "XML does not DO anything… It was created to structure, store and to send information. It [XML] is pure information wrapped in tags" (*Introduction to XML*). XML has no fixed tags; instead, in XML users define their own tags. User-defined XML tags allow the exchange of data between applications and systems that are not compatible. Castro (2001) states, "XML is a language for creating other languages. You can use XML to design your own custom markup language and then use that language to format your documents" (p. 13). This makes XML the perfect language for technical metadata storage because new applications can be written for the existing technical metadata, rather than rewriting technical metadata for the existing applications.

To have this degree of flexibility, XML documents must be formatted to specific criteria because data that will be rendered by numerous applications must have a uniform structure. The basic rules for writing XML are inclusion of a root element, closing tags, properly nested elements, case sensitivity, and enclosure of values in quotation marks (Castro, 2001, p. 23).

```
<?xml version="1.0" standalone="no"?>

<picture_database>
<title>Declaration of Independence</title>
<scan filename="Dec_of_Ind.jpg"/>
&DOI_descript;
</picture_database>
```

Figure 2. XML document

Figure 2 illustrates a simple XML document. Note that the root element opens with <picture_database> and closes with </picture_database>, and the <title> tag is properly nested within the root element. Since XML is case sensitive, "Declaration of Independence" is not the same as "declaration of independence". Additionally, the JPEG image "Dec_of_Ind.jpg"/> is a value; therefore, it is enclosed in quotation marks. The line &DOI_descript; is the name of an external Document Type Definition (DTD) that refers to a source outside the document to obtain a text description of the Declaration of Independence.

To make XML cross platform developers use another language, called extensible stylesheet language (XSL), to define the XML documents content and page attributes. W3 Schools defines XSL as a language that is based on XML standards (*XSL Languages*). XSL consists of three separate languages: extensible stylesheet language transformation (XLST), XML path language (Xpath), and XSL formatting objects (XSL-FO). XLST is a language that transforms XML documents into HTML or XHTML to enable rendering by browsers. Xpath locates data in an XML document, and uses expressions and functions to define criteria for data selection. XSL-FO is a formatting language that defines document content and page attributes.

To be useful, metadata not only needs to be read by applications, but also understood. Resource Description Framework (RDF) is a language that enables the exchange and processing of metadata by different applications. "RDF provides a common framework for expressing this information [metadata] so it can be exchanged between applications without loss of meaning" (W3C, *RDF Primer*). RDF defines the rules for the structure labels. Similar to sentences in English, RDF consists of a subject, a predicate, and an object. The subject describes what is being labeled, the predicate describes what the label contains, and the object describes the value of the contents (Adobe, *White Paper*, 2004, p. 11). Table 1 draws a comparison between a label for a photo written in English and a label written in RDF expressed in XML. Table 1 points out the core purpose of RDF—to take data that can be rendered by humans and convert it to data that can be rendered by machines.

Table 1. Comparison of human and machine rendered labels

English Language	RDF expressed in XML
Label of Photo	<rdf:Description=Photo>
Subject, Sue	<Subject>Sue</Subject>
Predicate, Shutter Speed	<ExposureTime>ShutterSpeed</ExposureTime>
Object, 1/80 second	<Value>1/80 second</Value>
	</rdf:Description=Photo>

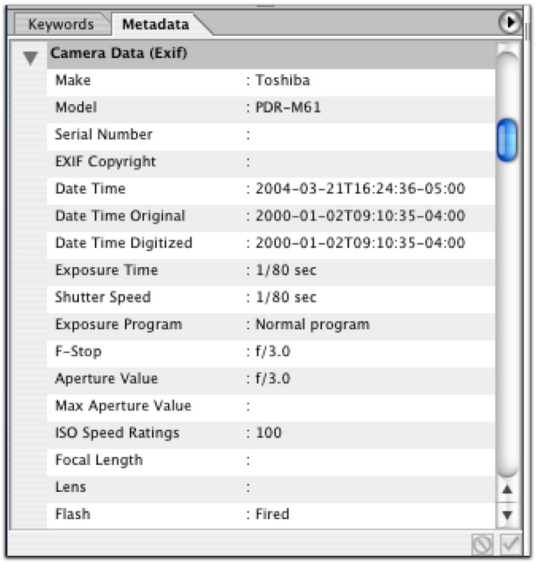
Adobe XMP. Supporting the XML standard, Adobe Systems uses a metadata language Adobe calls XMP (Extensible Metadata Platform). XMP is a packet of data consisting of XML statements, structured as RDF labels, embedded in an image file. "It

[XMP] is a labeling technology that lets users embed data about a file in the file itself, not just Photoshop TIFF and JPEG images" (Controlled Vocabulary.com). Adobe (2004) describes XMP as "a framework for adding machine readable labels semantic content to application files, databases and content repositories" (White Paper, p. 9). To appreciate Adobe's goal in adding semantic content to image files, W3C's (2001) description of the power of the semantic web can be utilized:

> The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URLs for naming.

According to Adobe (2004) XMP consists of three components: XMP framework, XMP schemas, and XMP packet technology (White Paper, p. 10). Adobe (2004) describes XMP framework as "a formal way of rendering into a label the commonsensical way of describing something" (White Paper, p. 11). XMP schemas are defined by Adobe (2004) as "The RDF rules that specify the composition of label into a sequence of XML statements structured as a triple of data called resource, property, value or alternatively called subject, predicate, object. The XMP packet is the binary structure that contains the label embedded in the document" (White Paper, p. 12).

Adobe (2004) asserts that XMP is designed according to RDF label structures and using XML expressions, thereby making it machine readable metadata (White Paper, p. 10). The CS version of Adobe Photoshop allows the user to view seventy fields of technical metadata. Figure 3 shows a portion of the viewable fields of metadata for image files along with corresponding XML code. Adobe (2004) cites the importance of XMP saying "a content composed for newspaper broadsheet will be repurposed for distribution via server to a Web browser on a PC, to a PDA, a cell phone, collected into an archive, and perhaps even pressed into a CD" (White Paper, p. 15).




```
<rdf:Description rdf:about='uuid:4edc3544-7ce4-11d8-8f3a-b6c8aefec736'
        xmlns:exif='http://ns.adobe.com/exif/1.0/'>
        <exif:ExposureTime>1/80</exif:ExposureTime>
        <exif:FNumber>302/100</exif:FNumber>
        <exif:ExposureProgram>2</exif:ExposureProgram>
        <exif:ExifVersion>0210</exif:ExifVersion>
        <exif:DateTimeOriginal>2000-01-02T09:10:35-04:00</exif:DateTimeOriginal>
        <exif:DateTimeDigitized>2000-01-02T09:10:35-04:00</exif:DateTimeDigitized>
        <exif:ShutterSpeedValue>633/100</exif:ShutterSpeedValue>
        <exif:ApertureValue>319/100</exif:ApertureValue>
        <exif:FlashpixVersion>0100</exif:FlashpixVersion>
        <exif:PixelXDimension>1792</exif:PixelXDimension>
        <exif:PixelYDimension>1200</exif:PixelYDimension>
        <exif:FileSource>3</exif:FileSource>
        <exif:Flash rdf:parseType='Resource'>
        <exif:Fired>True</exif:Fired>
        <exif:Return>0</exif:Return>
        </exif:Flash>
        <exif:ISOSpeedRatings>
        <rdf:Seq>
        <rdf:li>100</rdf:li>
        </rdf:Seq>
        </exif:ISOSpeedRatings>
 </rdf:Description>
```

Figure 3. Photoshop technical metadata fields and XML code.

Photo Mechanic. Photoshop is not the only application that uses an XML based graphic user interface (GUI) that allows users to enter and edit metadata. *Photo Mechanic* is a software package that is widely used in the newspaper industry to preview, edit, repurpose, and batch process images. *Photo Mechanic* allows viewers to tag thumbnail versions of the digital image and compare two-up displays for final selection. This quick visual selection process is important in a deadline driven business like newspapers. Additionally, *Photo Mechanic* allows users to tag images using International Press Telecommunications Council (IPTC) compliant descriptive metadata fields. *Photo Mechanic* also has the ability to extract technical metadata from images, and add it as a variable to IPTC fields. This has the advantage of including technical metadata with the image file, but only as a subset to the descriptive metadata field; therefore, this eliminates the ability of technical metadata to be used in any subsequent machine processing. Another disadvantage is that the inclusion of technical metadata is left to the user's discretion, so there is no standard of technical metadata inclusion. Despite these disadvantages in recording technical metadata, *Photo Mechanic* is popular in the newspaper industry because of its excellent graphic user interface, which allows quick editing (Camera Bits, 2004).

Capturing Technical Metadata

To capture and embed technical metadata in image files so it can be retrieved on any device requires the use of a standard encoding scheme. Currently, high-end digital

cameras capture technical metadata through the use of the Exchangeable Image File (Exif) format. Exif is a schema to embed technical metadata in JPEG, TIFF, and Camera Raw files. The purpose of Exif is to incorporate a uniform set of metadata into a digitally captured document that can be interpreted by any device used for image rendering. According to Kodak (2002) "Exif is an international specification that lets imaging companies encode metadata information into the headers or application segments of a JPEG file." The Japan Electronics and Information Technology Industries Association (2002) notes the versatility of Exif stating that it "specifies the formats to be used for images, sound and tags in digital still cameras and in other systems handling the image and sound files recorded by digital still cameras." Exif specifications are widely used in the imaging industry to record technical metadata, and Exif specifications provide what is probably the most common way to include technical metadata in JPEG files. Exif is not only a still digital metadata capture standard, but it is also a metadata standard for almost all types of digital recording. The strength of the Exif specification is that the information embedded in the image file can be read by any application supporting JPEG, TIFF, or Camera Raw files. Technical metadata including, but not limited to shutter speed, aperture, date and time of capture, and GPS location of capture is being included in the latest version of Exif specification (p. 1).

There are two criteria concerning the inclusion of technical metadata in image files—how to include data and what data to include. The Research Libraries Group Automatic Exposure–Technical Metadata initiative maintains a dialog with high-end camera manufacturers on how they capture technical metadata and identifies emerging

technologies "for the harvesting of technical metadata" (Günter & Dale, 2004, p. 260). The RLG sponsored initiative is exploring ways to include technical metadata into digital images in conformance with proposed National Information Standards Organization (NISO) Data Dictionary standards. Furthermore, the Automatic Exposure–Technical Metadata initiative has surveyed cultural institutions to determine the type of technical metadata that they want to be included in image files (Günter & Dale, 2004, p. 261).

The ability to capture technical metadata has been available in image file formats for over a decade; however, the use of metadata in image processing has not been widespread. The slow incorporation of technical metadata into image processing was due to the lack of universal technical metadata standards and the absence of a common encoding scheme. Kodak comments "for the full benefits of picture metadata to be realized across the open system, imaging applications first need a universal means to *define* metadata, and a means to *create*, *access*, *update and save* metadata associated with pictures" (Kodak, *Technology Highlights, Metadata*).

Technical Metadata Standards

Several organizations are working to meet the need for metadata standards in digital imaging. These include the International Press Telecommunications Council (IPTC), the Dublin Core Metadata Initiative (DCMI), the Idealliance Digital Image Submission Criteria (DISC), and the National Information Standards Organization (NISO). The following section discusses the objectives and standards established by each of these organizations.

IPTC Standards. The IPTC in conjunction with the Newspaper Association of America (NAA) has established an evolving set of metadata standards to describe what type of metadata to include in image files intended for the newspaper industry (Controlled Vocabulary.com, *The IPTC-NAA standard*). The primary purpose of the IPTC standard is to include descriptive metadata that establishes image ownership, caption, and location for image retrieval from databases. IPTC descriptive metadata standards are widely used in the newspaper industry, and they are supported by major news organizations like the Associated Press.

DCMI Standards. The DCMI consists of working groups that address the development of user specific metadata sets that can be distributed across domains. Table 2 shows the fifteen fields of metadata that are included in the initiative (DCMI, 2003).

Table 2. Elements of Dublin Core Metadata Initiative metadata fields

Title	Contributor	Reference Source
Creator	Date	Language
Subject	Resource Type	Relation to Reference
Description	Format	Scope of Coverage
Publisher	Resource Identifier	Rights Management

In addition to the named fields, cross references are provided for retrieval from Internet resources. While both the IPTC and DCMI have greatly enhanced digital image metadata, their primary purpose is that of including descriptive metadata.

DISC Standards. One group that is addressing technical specifications of digital image capture is Digital Image Submission Criteria. DISC, a working group of Idealliance, has established image quality and descriptive metadata standards for use by photographers and illustrators (Idealliance, *DISC Specifications*, 2003). DISC specifies digital camera settings based on the halftone quality needed for a specific reproduction process. DISC standards specify image capture settings by addressing, pixel resolution, color space, camera quality setting, and compression ratio. Figure 4 shows the DISC standards for digital image capture (Idealliance, *DISC Specifications*, 2003).

DISC Category A - 150 line Screen Reproduction

Camera/Scanner Pixel Resolution	Camera Megapixel Rating	Color Space Setting	Image Quality* Setting	Compressed File Size	Uncompressed File Sze	Max. Print Size (300ppi or 150lpi)
4500 x3000	13.5	AdobeRGB1998	High Quality(8)	1.67MB	38.32MB	15.0 x 10.0
4064 x 2704	11.0	AdobeRGB1998	High Quality(8)	1.36MB	31.20MB	13.5 x 9.0
3008 x 2000	6.0	AdobeRGB1998	High Quality(8)	0.74MB	17.08MB	10.0 x 6.7
2560 x 1950	5.0	AdobeRGB1998	High Quality(8)	0.62MB	14.17MB	8.5 x 6.5
2448 x 1632	4.0	AdobeRGB1998	High Quality(8)	0.49MB	11.34MB	8.2 x 5.4
2000 x 1500	3.0	AdobeRGB1998	High Quality(8)	0.37MB	8.52MB	6.7 x 5.0
1600 x 1200	1.9	AdobeRGB1998	High Quality(8)	0.24MB	5.45MB	5.3 x 4.0
1280 x 1240	1.3	AdobeRGB1998	High Quality(8)	0.16MB	3.72MB	3.4 x 3.4
640 x 480	0.3	AdobeRGB1998	High Quality(8)	0.04MB	0.87MB	2.1 x 1.6

DISC Category B - 125 (133) line Screen Reproduction

Camera/Scanner Pixel Resolution	Camera Megapixel Rating	Color Space Setting	Image Quality* Setting	Compressed File Size	Uncompressed File Sze	Max. Print Size (250ppi or 125lpi)
4500 x3000	13.5	AdobeRGB1998	High Quality(8)	1.67MB	38.32MB	18.0 x 12.0
4064 x 2704	11.0	AdobeRGB1998	High Quality(8)	1.36MB	31.20MB	16.3 x 10.8
3008 x 2000	6.0	AdobeRGB1998	High Quality(8)	0.74MB	17.08MB	12.0 x 8.0
2560 x 1950	5.0	AdobeRGB1998	High Quality(8)	0.62MB	14.17MB	12.2 x 7.8
2448 x 1632	4.0	AdobeRGB1998	High Quality(8)	0.49MB	11.34MB	9.8 x 6.5
2000 x 1500	3.0	AdobeRGB1998	High Quality(8)	0.37MB	8.52MB	8.0 x 6.0
1600 x 1200	1.9	AdobeRGB1998	High Quality(8)	0.24MB	5.45MB	6.4 x 4.8
1280 x 1240	1.3	AdobeRGB1998	High Quality(8)	0.16MB	3.72MB	5.1 x 4.1
640 x 480	0.3	AdobeRGB1998	High Quality(8)	0.04MB	0.87MB	2.6 x 1.9

DISC Category C - 100 line Screen Reproduction

Camera/Scanner Pixel Resolution	Camera Megapixel Rating	Color Space Setting	Image Quality* Setting	Compressed File Size	Uncompressed File Sze	Max. Print Size (200ppi or 100lpi)
4500 x3000	13.5	AdobeRGB1998	High Quality(8)	1.67MB	38.32MB	22.5 x 15.0
4064 x 2704	11.0	AdobeRGB1998	High Quality(8)	1.36MB	31.20MB	20.3 x 13.5
3008 x 2000	6.0	AdobeRGB1998	High Quality(8)	0.74MB	17.08MB	15.0 x 10.0
2560 x 1950	5.0	AdobeRGB1998	High Quality(8)	0.62MB	14.17MB	12.8 x 9.8
2448 x 1632	4.0	AdobeRGB1998	High Quality(8)	0.49MB	11.34MB	12.2 x 8.2
2000 x 1500	3.0	AdobeRGB1998	High Quality(8)	0.37MB	8.52MB	10.0 x 7.5
1600 x 1200	1.9	AdobeRGB1998	High Quality(8)	0.24MB	5.45MB	8.0 x 6.0
1280 x 1240	1.3	AdobeRGB1998	High Quality(8)	0.16MB	3.72MB	6.4 x 5.1
640 x 480	0.3	AdobeRGB1998	High Quality(8)	0.04MB	0.87MB	3.2 x 2.1

*Photoshop High Quality JPEG compression level 8

Figure 4. DISC standards for digital image capture.

Also, DISC has established thirteen fields of descriptive metadata to be included with a digital image. Furthermore, DISC has created an XMP based plug-in that allows Adobe *Photoshop CS* to read and write descriptive metadata fields (Idealliance, *DISC XMP Custom Panels*, 2004). While DISC addresses the issue of technical specifications for digital capture, it does not include technical information in its metadata fields.

DIG35 Standards. Digital Imaging Group known as DIG35 has created standards for still digital image metadata. The stated goal of DIG35 (2000) is "to define a standard set of metadata for digital images that will improve the semantic interoperability between devices, services, and software" (p. 1). In other words DIG35 (2000) intends to design a set of metadata that is interchangeable, extensible and scaleable, image file format independent, consistent, and Internet ready (p. 2). DIG35 (2000) has proposed several categories of metadata to be captured, including basic image parameter, image creation, content description, intellectual property rights, and fundamental attributes. Of these categories image creation metadata addresses the issue of standards for technical metadata. Image creation metadata includes information about capture, capture device, software used, optical parameters, and device characterization.

NISO Standards. Murphy (2004) describes the National Standards Information Organization (NISO) as "a non-profit association accredited by the American National Standards Institute (ANSI), which identifies, develops, maintains, and publishes technical standards to manage information in the digital environment" (p. 60). NISO standards cover the areas of information-related needs, including retrieval, repurposing, storage, metadata, and preservation. NISO (2004) comments "cultural institutions had been focusing primarily on defining descriptive metadata for the purpose of discovery and identification, and that comparatively little work had been done to codify technical attributes of digital images and their production" (p. 2).

To create a uniform metadata standard, NISO has been working on a set of technical metadata standards in its *Data Dictionary—Technical Metadata for Still Digital Images*. The purpose of the *Data Dictionary* (2004) is to "to define a standard set of metadata elements for digital images" (NISO, p. 8). The editors of the *Data Dictionary* note, "Technical metadata have been identified to 'anchor' meaningful attributes of image quality that can be measured objectives, such as detail, tone, color, and size" (NISO, 2004, p.8). *The Data Dictionary* addresses the need to establish firm categories of technical metadata for use with image files because NISO considers uniform mapping of technical metadata essential if metadata is to be automatically collected.

The *Data Dictionary* clearly differentiates between descriptive and technical metadata. Descriptive metadata is not about the image file itself; it is about the conditions surrounding the image capture, such as the photographer, the location, and ownership of the image (NISO, 2004, p. 9). The realm of Data Dictionary is to describe the "basic image parameters…[that] are fundamental to the reconstruction of the digital file as a viewable image on an electronically interfaced displays"(NISO, 2004, p. 13). Therefore, technical metadata is specific to the image, such as color space, compression scheme, embedded ICC profile, and encoding scheme (NISO, 2004, p. 13). An important goal of the *Data Dictionary* is to capture technical metadata for archival purposes. The *Data Dictionary* observes, "none of the current vehicles for technical metadata were explicitly created with digital preservation in mind" (NISO, 2004, p.262). It should be noted that image authenticity and integrity are not part of these specifications (NISO, 2004, p. 8).

NISO (2004) defines four categories of technical metadata to be included with image files: basic image parameters, image creation, image performance assessment, and change history. Basic image parameters are those attributes that are considered essential to be able to read and display the image file. Image creation metadata is data surrounding the image capture and capture system used (NISO, 2004, p. 25). Describing image performance assessment, NISO (2004) comments image performance assessment "has both a present and future context: these elements serve as metrics to access the accuracy of output (today's use), and to access the accuracy of preservation techniques, particularly migration (future use)" (p. 35). NISO (2004) points out "The metadata blocks in Change History are used to document the source, systems, and settings used in all subsequent digital-to-digital operations" (p. 55). Table 3 compares NISO mandatory technical metadata with TIFF, DIG35, and Exif 2.2.

Table 3. The quantitative result of mapping all four sections of the NISO Z39.87 data dictionary to TIFF, DIG 35, and Exif

	NISO Z39.87 Complete	**TIFF 6.0 / TIFF EP Mapping Elements**	**DIG 35 1.1 Mapping Elements**	**Exif 2.2 Mapping Elements**
Basic Image Parameters	30	16	8	13
Image Creation	38	9	34	19
Image performance Assessment	36	22	7	16
Change History	7	2	6	1
Total	111	49	55	49

Günter & Dale, 2004, p. 262. Copyright 2004 by The Society for Imaging Science and Technology. Adapted with permission.

There are obvious differences in the type of technical metadata included in various standards. TIFF and Exif standards lack several categories of change history, and DIG35 captured only seven of a possible thirty-six Image Performance Assessment fields. Even though NISO is the most extensive standard, it is important to note that some of the NISO fields can only be included by manual intervention.

Each of the organizations serves a different purpose. The IPTC and DCMI address the inclusion of descriptive metadata so image files can be cataloged and shared. DISC adds the additional criteria of image capture standards based on intended reproduction. DIG35 and NISO address the need for a uniform set of technical metadata that can be automatically captured and parsed by image processing devices. Table 4 summarizes the inclusion of metadata by each of these organizations.

Table 4. Overview of metadata inclusion by organization

	IPTC	DCMI	DISC	DIG35	NISO
Descriptive Metadata (User entry)	X	X	X		
Specification of capture settings			X		
Technical metadata (User entry)				X	X
Technical metadata (Automatic entry)				X	X

Technical Metadata for the Newspaper Industry

McDowell contends that the purpose of technical metadata is to help reproduce or enhance an image; therefore, any information not used for this purpose is not technical metadata. To illustrate his point, McDowell draws a distinction between images intended for consumer use and images intended for the graphic arts. He states that consumer photographic processes use algorithms to render a pleasing image based on an average input. However, in the graphic arts, images are often individually adjusted for a specific purpose. For example an image for advertising might contain several elements. One of these elements, the product being sold, is selectively enhanced. This advertising image, prepared for a specific purpose, requires considerable human intervention and subjective decision-making (McDowell, D., personal communication, August 19, 2004).

Likewise, a distinction can be drawn between editorial images and advertising images. McDowell illustrates his point by comparing editorial images to the advertising images in magazines. The advertising images are usually better because more subjective human intervention was used to create them. Spending more time on advertising images is justified because the primary purpose of advertising images is to generate product revenue (McDowell, D., personal communication, August 19, 2004).

The extent of subjective human intervention also brings up the question of where the original image begins. If the originally captured image serves no purpose other than to provide raw material for an illustration, is the data included at capture technical metadata? Frey (1999) supports this, pointing out that in archival collections of fine art photography, the artist often spends considerable darkroom time altering the negative;

therefore, the print should be treated as the original to be scanned into the collection.

To answer the question of the usefulness of technical metadata to newspapers, one must understand that the purpose of newspaper photography is to display editorial content. Therefore, newspapers have a vested interest in maintaining the integrity of the image at the point of capture. As long as the editorial integrity of a newspaper image is preserved, image reproduction quality may be decided by viewer preference. Preserving the editorial integrity of an image means portraying the events depicted without bias. However, there is a fine line between alterations that are considered preferential and alterations that alter image content. In the past, several major news organizations have unintentionally changed the editorial meaning of an image through a seemingly minor image adjustment. Nevertheless, once the editorial integrity of a newspaper image is ensured, the image should be pleasing to the viewer.

Viewer preference for image quality fits well with the Hunt's preferred color reproduction objective, which is normally associated with consumer photography. Hunt (2004) defines preferred color reproduction "as reproduction in which the colours depart from the equality of appearance to those of the original, either absolutely or relative to white, in order to give a more pleasing result to the viewer" (p. 174). Importantly, preferred color reproduction lends itself to automation because consumer expectations for output of pleasing images are known. Given the spectrum of reproduction objectives, newspaper images are reasonably close to preferred color reproduction; therefore, they are good candidates for automation.

Chapter Summary

Metadata serves several purposes. Metadata can be administrative, assisting in the management of digital resources. It can be descriptive, identifying and categorizing a digital resource. Metadata can verify the integrity of a digital resource, and it can supply information about the conditions of image capture. Technical metadata specifically records capture device settings and the conditions surrounding the capture.

From their inception, file formats like TIFF and JPEG have been able to include technical metadata at image capture. The Exif schema has created a uniform set of technical metadata to be included with digitally captured images.

To include metadata after image capture, corporations such as Adobe are actively supporting the use of XML as a uniform open standard for the inclusion of technical metadata. Lastly, organizations such as DIG35 and NISO are taking a leadership role in determining what type of technical metadata to include with image capture. The stage is being set for the use of technical metadata in the digital images used by newspapers.

Chapter 3

Research Questions

In addition to images, digital files include a significant amount of associated information, called technical metadata, which describes the tools, the settings, and the conditions surrounding digital capture. Due to the time sensitive nature of newspaper publishing, news images are increasingly being captured digitally. It is essential that newspapers improve the efficiency of their digital workflows, while maintaining traceability of the digital file to its source of capture. The efficient reproduction as well as the editorial integrity of these image files is essential.

To meet these objectives using technical metadata, the first step is to determine the fields of technical metadata that are relevant to the newspaper industry. To date, the NISO *Data Dictionary* describes the most comprehensive set of technical metadata standards, listing 111 fields of technical metadata. Consequently, this research will evaluate the technical metadata needs of the newspaper industry in two key areas based on the NISO *Data Dictionary* standard. It will:

- Examine how newspapers currently use technical metadata in their digital imaging workflow.
- Point out areas where technical metadata could improve the newspaper digital imaging workflow.

Chapter 4

Methodology

Data Collection Plan

Five newspaper organizations were selected for this study. The organizations studied were *USA Today*, *The Chicago Tribune*, *The Washington Post*, *The New York Times*, and *The Rochester Democrat & Chronicle*. The first four organizations were chosen because they were representative of the largest newspaper organizations in the United States. In addition, the Rochester Institute of Technology had established ties with each organization, which facilitated the completion of interviews within the time allotted for research. *The Rochester Democrat & Chronicle* was added as an initial case study. The plan was to access the effectiveness of the interview technique using the initial case study, and then, make necessary adjustments for future studies. However, on its own, the *Democrat & Chronicle* was considered to be important, because it was the first major metropolitan newspaper in the United Stated to convert to an all digital image workflow.

To obtain sufficient information to draw conclusions for both research questions, the participants in each case study were asked a number of structured questions (See Appendix A) about their organization's digital workflow and their use of technical

metadata. To allow additional input, each interview provided ample time for open-ended responses.

Additionally, a questionnaire (See Appendix C) was used to assess the perceived importance of different types of technical metadata. Case studies were important to this research because first hand observation of each organization's digital workflow was an important component of the study. Furthermore, on-the-spot questioning provided the flexibility to account for each organization's unique description of technical metadata.

Data Analysis

The first research question sought to understand if technical metadata would improve the newspaper digital imaging workflow, and the type that would do so. The second research question sought to determine the type of technical metadata essential to preserve the integrity of newspaper digital images. Questions 1 through 3 prompted the user to describe the scope of the organization's digital workflow and its need for image quality. Questions 4 through 6 asked the user to assess the importance of using technical metadata in the capture and the adjustment of images. As may be seen in Appendix A, questions 8 through 11 asked users specific questions about their current technical metadata practices based on NISO standards. Questions 12 and 13 inquired about the preservation of technical metadata as digital files were repurposed or converted to another file format. Additionally, a series of survey questions Appendix B) was included to provide a user defined assessment of the importance of specific technical metadata fields.

This study used NISO definitions to set specific boundaries to improve the quality and the consistency of responses, since this standard, to date, represents the most comprehensive set of technical metadata in existence. Obviously, these boundaries limited the nature of the research, causing it to be more deductive than inductive. However, as Miles and Huberman (1994) pointed out "better research happens when you make your framework and associated choices of research questions, cases, samples, and instrumentation explicit, rather than claiming inductive purity" (p. 23). Moreover, the use of an explicit survey instrument allowed better cross-case comparison.

To assist in designing interview questions a conceptual workflow was constructed (See Appendix D). This flowchart served the purpose of creating a visual framework that illustrated critical handoffs in the workflow. Based on this hypothetical workflow, the digital imaging process was divided in the following manner: Image capture, Image Editing, Image Storage, and Image Output. The flowchart represented generic categories, and in some organizations the actual workflow differed considerably; however, the processes of capture, editing, storage, and output were common in all organizational structures.

To compare the responses from different organizations, a matrix was constructed using information gathered from the case studies (See table 5 on page 59). The purpose of this matrix was to show both within-case and cross-case responses. Once the data was entered into the matrix, relationships emerged, and based on these relationships, the data was sorted and clustered into defined patterns. As Miles and Huberman (1994) comment,

cross-case analysis is necessary to be able to better generalize results, and to "deepen understanding and explanation" (p.173).

To analyze Likert scale[1] data from the Case Study Questionnaire, the assumption was made that the responses represented interval data. Undoubtedly, with any qualitative measurement, the assurance of a true interval scale could not be guaranteed. However, the lack of a true interval scale in Likert-style responses did not seriously affect a meaningful comparison of responses. As Jacquard and Wan (1996) assert, "for many statistical tests, rather severe departures [from intervalness] (stet) do not seem to affect Type I and Type II errors dramatically" (p. 4). Finally, information gathered from the questionnaire was blended with interview responses and studied for trends as well as novel ideas.

Endnotes Chapter Four

[1]The Likert scale is a four or five step ordinal scale that asks the respondent to assign a level of importance or agreement with a particular question. The following scale was used in this research project.

5 = Very important
4 = Important
3 = Moderately important
2 = Of little importance
1 = Unimportant

Chapter 5

Summary of Results

The purpose of this research project was to examine the use of technical metadata in the newspaper industry. Chapter Four outlined the data collection and analysis plans. Through the use of a survey instrument and in-person interviews, this research sought to answer two questions about the use of technical metadata by newspaper organizations:

- Examine how newspapers currently use technical metadata in their digital imaging workflow.
- Point out areas where technical metadata could improve the newspaper digital imaging workflow.

This chapter summarizes the results from both the interviews and the questionnaires (Refer to Appendix B for a list of interview questions and Appendix C for the complete case study questionnaire). Each summary starts with a broad stroke description of the organization followed by a description of the organization's digital imaging workflow. Next is a summary of the responses to the interviewer's questions about technical metadata, and a summary of open-ended responses. At the conclusion of the summaries, there is a matrix of questionnaire responses.

Case Study One

Introduction

USA Today the national, and international, newspaper of the Gannett Company, published at forty print sites worldwide. The total daily circulation of *USA Today* is in excess of two million copies. Printing at forty sites creates a two-tier distribution scheme in which *USA Today* headquarters in McLean, Virginia, transmits pages electronically to customer print sites, which in turn, distribute printed newspapers to end consumers. In addition to printed products, *USA Today* maintains a news website at www.usat.com. *USA Today* acquires hundreds of thousands of photos yearly from its photographers, news service providers, and through broadcast captures. Out of this huge quantity of images, it is estimated that only about five percent are saved, and even fewer are published. Steve Terrillion, Director of Prepress operations for *USA Today* comments, "We might take five hundred photos of the president and only use one."

Digital Imaging Workflow

All photos taken by *USA Today* staff photographers are captured digitally. Staff photographers capture images as Camera Raw files. *USA Today* is acutely aware that minor changes in a digital image may alter editorial content; therefore, the organization wanted a digital image file that had not been processed through camera software. Terrillion declared that because of image adjustment, some newspapers did not show all the debris of the Challenger explosion and that changed the event. No specific

sample per inch (spi) requirement was indicated, but *USA Today* advertising specifications require that digital images be between 175 and 250 spi. However, Terrillion pointed out "we won't select or not select an image because of some attributes of the image. It's [about] news content."

Photographers download digital files to workstations using *Photoshop*, for the purpose of previewing and tagging images. Subsequently, these files are saved as RGB JPEG to a *MerlinOne* digital asset management system. *Merlin One* is a suite of software products that manages collections of images and allows text searching, using IPTC standards-based metadata. In *Merlin*, a handoff occurs from photographer to photo editor. Working strictly in RGB, photo editors sharpen images and crop for content. The RGB file is then placed in proper position on the electronic page. In general, editors do not adjust the color of images; that is left to the production department. Terrillion declares that there is a clear separation in the type of image adjustment done between departments. The editorial department adjusts images for content, and the production department adjusts images for press.

In the production area, digital images are first converted to CMYK, and then visually adjusted. At this time, *USA Today* is not using a color-managed workflow[1]; it is relying on the expertise of its staff to adjust images. After the CMYK image is adjusted, it is returned to the queue, where it replaces the original RGB image that was placed on the page. Currently, the editorial department is saving both the RGB and the CMYK version of each image in *Merlin*. Once the electronic page is complete, color separated G-4 TIFF files are sent to print sites. Sending G-4 TIFF files to print sites eliminates

subsequent file processing, thus reducing onsite errors. In addition to archiving images, *USA Today* archives complete pages as PDFs in various places throughout the organization. These PDF files serve numerous purposes, such as proofs for advertisers or reprints for customers.

Metadata—Response to Questions

Images are tagged with descriptive metadata according to IPTC descriptive metadata standards. The organization does use some technical metadata that falls under the NISO heading of "Basic Image Parameters", such as file size and color space, and it is mindful of other metadata such as degree of compression. However, none of this metadata is used in an automated fashion. Photographers must intentionally enter "Image Creation Metadata" by adding date, location, and the creator of the image. The *Merlin* archiving system does automatically create a unique identifier for images. Interestingly, the use of Camera Raw image files by *USA Today* photographers causes photographers to use Image Creation Metadata to assist the visual adjustment of images. Miles Weissman, Prepress Systems Administrator, commented that in the early stages of conversion to digital photography, Image Performance Assessment tools like test targets were used due to the limitations of the equipment, but these tools are not currently being used. Change History is kept by saving versions of images. As Terrillion pointed out, since a detailed change history would require considerable human intervention, versioning is probably the best method of tracking changes in a busy newspaper environment. When asked whether technical metadata would be kept intact during a migration to a new file format,

Weissman remarked that it would depend on the application that performed the conversion.

Metadata—Open-Ended Responses

Asked to speculate on the usefulness of technical metadata, Terrillion remarked that the more information there is about an image the better, but there may be limitations due to the digital overhead of carrying technical metadata in an image file. He also noted that the interplay of the large systems that *USA Today* uses to manage photos determined how much useful technical metadata an image could have. Additionally, Terrillion noted that news deadlines determine what a newspaper will or will not do in an image archive. He added that the inclusion of technical metadata needed to be automated by the application used to ensure that it is kept on a regular basis. A final concern was that the amount of metadata in an image file could overwhelm human users. Speculating on the potential positive aspects of technical metadata, Terrillion thought that technical metadata could be a good tool for decision makers to improve production methods and to recommend equipment expenditures.

Terrillion spoke extensively about the use of metadata for advertising images. *USA Today* creates thousands of ads yearly from customer-supplied images. Areas deemed important to track, though not strictly technical metadata, are size, billing, press run date, and multimedia functionality. Terrillion declared, "An ad might have multimedia functionality built in that you don't see on the printed page, but when it goes out in electronic format [it] might have some of those options." Advertiser access to

information was considered to be an area of ongoing research. It would be desirable to transmit a customer's ad on a completed news page containing metadata about billing and insertion. However, controlling customer access to metadata on a newspaper page containing several different elements is a challenge, because each customer would have access to the metadata contained in every advertisement.

Terrillion brought up the point that the same metadata can be interpreted differently depending the context that it is used in. For one application certain metadata may be considered technical according to NISO standards, while in another application it may be considered descriptive according to standards like AdsML. AdsML is an international standard for digital advertising exchange supported by Ifra[2] and the NAA. "The AdsML Standard is an XML-based…open standard that will manage the digital exchange of information between all parties involved in the advertising workflow, from concept to publication to billing" (*AdsML Consortium*, 2004).

Case Study Two

Introduction

The *Chicago Tribune* has a daily circulation of approximately 790,000 subscribers, increasing to one million for the Sunday edition. The Tribune's primary distribution area is metro Chicago and the surrounding fifty-mile area. The *Chicago Tribune* is part of the Tribune Company which publishes several other newspapers and magazines, most notably, the *Los Angles Times*. In addition to newspapers, the Tribune Company owns broadcast media such as television stations WGN in Chicago and WPIX

in New York. The organization maintains an interactive *Chicago Tribune* website. In addition to its news outlets, the Tribune Company owns the Chicago Cubs baseball team.

Currently, ninety percent of the images taken by *Chicago Tribune* photographers are taken digitally. The *Chicago Tribune* publishes about two hundred images per day. Kelly Hagen, Advertising Systems and Client Relations Manager, estimated that the Tribune captured approximately twenty images for each one published. There are future plans to install a system that will report the ratio between captured and published images. Hagen added that the *Tribune's* advertising department accepts about 10,000 images and digital elements yearly from customers. Additionally, the organization accepts approximately 15,000 complete digital advertisements, all in the form of PDF files.

Digital Imaging Workflow

Tribune photographers capture images as JPEG files, except for high-end Sunday magazine images that are captured as Camera Raw files. Associate Subject Editor, Rick Berry, commented that capturing Camera Raw files allowed greater ability to correct exposure after the shoot. All digital images are captured at a resolution of 200 pixels per inch (ppi). The *Tribune* does not use a color-managed workflow. Color evaluation is done on a proofing device that is calibrated to emulate the output of the press.

A *MerlinOne* archiving system is used for image storage. To initiate a photo shoot, the editor creates an assignment sheet in Merlin to be used by the photographer. This assignment sheet contains data such as place and time of the photo shoot. After the shoot, photographers caption images in *Photoshop* or *PhotoMechanic* according to IPTC

descriptive metadata standards, make minor tonal adjustments in Photoshop, and add them to the Merlin database. Image sharpening and color correction are left to the production department. Berry stressed that ethical considerations are important, and no adjustments are made past the point of making images look good in the newspaper. Thereafter, images are keyed to the assignment sheet in Merlin, which links the image with previously entered descriptive metadata.

The majority of news pages are paginated by the *Tribune's* CCI pagination system[3], which allows editors to place RGB images on the page for later conversion to CMYK. When images are published from CCI, the publication data is sent to Merlin where the high-resolution image is linked to the page. Published images are then reclassified to prevent being purged from the system. Currently, the *Tribune* has 600,000 RGB images archived in Merlin. Berry noted that CMYK images used in publication are rarely kept. If image is used again the original RGB file is adjusted and converted to CMYK.

Metadata—Response to Questions

The *Chicago Tribune* uses technical metadata that falls under the NISO classification of Basic Image Parameters, such as file size, compression, and color space. One significant difference from other organizations' operations is the *Tribune* uses ColorMatch RGB color space because it has a gamma that approximates that of print reproduction. Therefore, during the subsequent conversion to CMYK, darker areas of an image have less total ink coverage, which is important in high dot gain newspaper

printing. The use of ColorMatch RGB represents a more output-centric philosophy than that of other newspaper organizations. The Tribune also uses technical metadata regarding image resolution to automate the conversion of print images for the web.

Berry reported that the organization saves image creation metadata, or what it calls "exit data," to provide feedback about image reproduction. He admitted that image creation metadata could be useful for image adjustment technicians too; however, technicians would need a thorough understanding of photography. For technicians, Berry noted that "that the ultimate target… is a proofer that matches the press." Test targets are used to calibrate cameras and scanning devices, but the *Tribune's* workflow is not ICC color managed. Some of the organization's photographers use a Kodak gray card as an image performance tool. To verify neutral gray reproduction, the gray card is sometimes reproduced and checked for neutrality; however, target verification data is not stored. The *Tribune* does not keep a detailed change history, but saves versions of images. Berry pointed out that the originally captured RGB image, which they call an "outtake," is always saved to removable storage media such as a CD. In the future, it is planned to save these outtakes in the Merlin system, but it is doubtful if legacy files will be added. In the event of migration to a new image file format, it was thought that preservation of technical metadata would not be a priority.

Metadata—Open-Ended Responses

The *Chicago Tribune* actively shares its images with other Tribune publications. Additionally, the Tribune has a good working relationship with *Sports Illustrated*, which

occasionally publishes images taken by staff photographers. Member newspapers are able to log into the *Chicago Tribune's Merlin* database to search for images. Berry considers internal sharing a viable alternative to going to outside sources like the Associated Press. Obviously, in such a sharing scheme, the standardized use of descriptive metadata is important. Even though there is extensive sharing of images within the organization, there is no provision for outside access to images. The *Tribune* has a collection of thousands of glass plate negatives, and currently there is no plan to add these archival images to a database or share them. Upon further investigation, this researcher did find a collection of approximately 300 current and historical photographs for sale by a third party image management firm called International Published Photos Incorporated (IPPI). IPPI is an organization that provides a complete solution for organizations with large image archives to market their images and pages to retail consumers. Photographs are purchased through the *Tribune's* website and customer billing is conducted by IPPI.

Case Study Three

Introduction

The *Washington Post* has an approximate daily circulation of 780,000 subscribers, increasing to 1.1 million on Sundays. The *Washington Post* is a component of the Washington Post Company, a conglomerate consisting of newspapers, magazines (notably *Newsweek),* broadcast media, WPNI (Washington Post Newsweek Interactive),

and Kaplan Higher Education. Additionally, the Post has a longtime working relationship with Tribune's *Los Angeles Times*.

The Post maintains bureaus throughout the world to cover news events. Kevin Conner, Quality Assurance Manager for the Post estimated that about 32,000 digital images are inputted yearly into the organization's Merlin system. Although there was no exact count, Conner estimated that only about twenty percent of captured images were entered into *Merlin*. After capturing images, staff photographers return several selected images to the photo editor, or whoever assigned the shoot. After careful review, the editorial board decides which images to publish. The editorial department then communicates the intent of the image to the production department to ensure editorial integrity.

Digital Imaging Workflow

The Post does not use a color-managed workflow. Proofing devices are calibrated to emulate newsprint output. Staff photographers captured digital images as JPEG sRGB files; however, feature photos are occasionally captured as Camera Raw files. The minimum file requirement is 170ppi, but Connor mentioned that photographers frequently capture images at 300ppi to enable use in other publications. The assigning editor enters information about the photo shoot into the *Merlin* system, and the photographer inputs photos and caption information. Prior to entering an image into *Merlin*, photographers perform some image adjustment in Photoshop. Digital files are entered into the *Merlin* system as RGB, where they are later converted to CMYK by the

production department. Frequently, photographers add written instructions for the production department that helps to preserve the creative content of the image. The organization uses *Merlin* as its primary image archiving application, and editors use a CCI pagination system in conjunction with *Merlin* to place RGB images on news pages.

Tim Fitzsimons, Engraving Foreman, expressed concern that photographers sometimes altered the original image file in the *Merlin* database, thereby destroying the high-resolution image. Fitzsimons also spoke of the difficulty in reaching customer expectations when reproducing certain iridescent colors that sports teams use for their uniforms. Summing up the organization's digital image workflow Connor declared "we have to be careful with the word workflow because in the traditional sense the workflow applies to the daily publication. Period."

Metadata—Response to Questions

Examining the *Post's* use of technical metadata classified as Basic Image Parameters, the organization uses file type, file size, compression, and color space. One area of technical metadata the *Post* diligently tracks is the amount of image compression used. Connor spoke specifically about the relationship between compression and image quality. Staff photographers compress images to speed transmission, and at times, the subsequent reduction in image quality is at odds with the desired reproduction quality.

Both Connor and Fitzsimons thought that Image Creation Metadata could be useful for improving reproduction quality. Based on their interview responses, the organization actively pursues image quality improvement from a visual perspective.

Fitzsimons remarked that he has created film output from digital input to establish a common language to discuss image quality issues with photographers. Both Connor and Fitzsimons emphasized the importance of working with staff photographers during the conversion from analog to digital photography. A current area of discussion is the relationship between the on-screen view of an image and its reproduction in the newspaper. The organization does not currently use Image Performance Tools such as test targets. Connor pointed out the difficulty of setting up test targets at important and breaking news events. Nevertheless, Fitzsimons commented that test targets, such as a Macbeth ColorChecker, have been used in house to calibrate output.

As with other organizations, versioning is the primary means of keeping a record of the Change History of an image. Fitzsimons commented that there was some image adjustment that is done prior to inputting the image into *Merlin*, and there was no record of these adjustments other than the difference between the captured image and the image entered into *Merlin*. The issue of preserving technical metadata when migrating to another file format has not been discussed within the organization. Connor did not express strong feelings about the potential usefulness of technical metadata in the future. As far as Connor and Fitzsimons knew, *Merlin* did not provide complete access to technical metadata. However, this may be incorrect because they had not had a need to access technical metadata through the *Merlin* system. Fitzsimons brought up the point that staff photographers may already be using technical metadata to some extent without realizing it.

Metadata—Open-Ended Responses

The *Washington Post* is actively pursuing alternative channels of distribution for its news. One alternative channel of distribution is a website titled *Camera Works* that is linked to the main *Washington Post* website. *Camera Works* is a multimedia website consisting of images taken by both staff photographers and images submitted by outside sources. After selecting a storyline, customers are guided through an advertisement to either a slide show or a video presentation. According to website information, published newspaper images taken by staff photographers after 1996 can be purchased.

Another alternative channel of newspaper distribution is out of market sales. Fitzsimons mentioned that a *Post* reader in a Singapore hotel could print out a complete same day copy of the newspaper. This process involves sending a PDF file of the complete newspaper to a third party remote location. The service provider then digitally prints an on-demand copy of the newspaper. Typically, the digital edition is an 11 x 17" version of the actual newspaper.

Case Study Four

Introduction

The *New York Times* is distributed both nationally and internationally with its primary distribution area being New York City. The *Times* has a daily circulation of 1.1 million subscribers and a Sunday circulation of 1.6 million. The targeted advertising market distinguishes the national and local versions of the newspaper. In addition to the newspaper, the New York Times wire service distributes content to other publications,

and the New York Times Syndication Service distributes syndicated stories to publications nationwide. Tara House, Managing Director of Systems and Technology for the *Times*, comments "That [syndicated] content tends to be written by freelance journalists. It includes our *Book Review* section, our *Magazine* section, and our *Op-Ed* section." Additionally, the organization publishes a small (approximately eight page) international edition that is inserted into international newspapers. Furthermore, the New York Times Digital Group maintains a web-based publication. The New York Times Company also owns publications such as the *International Herald Tribune*, *The Boston Globe*, and sixteen other newspapers, as well as forty news websites.

Digital Imaging Workflow

House remarked that it was difficult to ascertain exactly how many digital images the Times captured, but the ratio of images captured to images published was greater than ten to one, and for important news events the ratio might be as high as one-hundred to one. The organization's digital workflow starts when a staff photographer is given a photo assignment. House notes that the technique used by photographers on assignment depends both on the subject matter and the means used to transmit the image to the newspaper. Rather than specifying a particular capture resolution, photographers adjust resolution based on the proposed end use of the image. Staff photographers sometimes capture images in Camera Raw format, but all image files are converted to JPEG before being entered into the image database. Similar to other large newspaper organizations, the *New York Times* uses a *MerlinOne* system to archive its image files. In addition to

assigned photos, staff photographers can, at their discretion, input additional images into Merlin. In Merlin, photographers use descriptive metadata fields to caption images for later identification.

The organization's library group archives images using descriptive metadata based on a set of 400 keywords; however, even with the addition of keywords, the original photographer's caption is maintained as a permanent record. House added that the organization does not use standard IPTC fields to add descriptive metadata. During the editing and page layout process, the image is usually kept in JPEG file format; however, the final image is saved as an EPS file. When queried as to why the Times preferred EPS files, House declared, "We don't have problems processing our pages, and EPS is easier to troubleshoot." Additionally, EPS files give the opportunity to include additional information with the image file.

The *New York Times* uses ICC color management to convert photos for publication; however, skilled technicians still make local image adjustments on a visual basis. House emphasized that the Times commitment to good journalism and the organization's vast resources allowed it to devote extra attention to image quality.

Metadata—Response to Questions

Like other large newspaper organizations, the *New York Times* digital workflow relies on technical metadata under the NISO classification of Basic Image Parameters, such as file type, file size, and image compression, to process images. House speculated that Image Creation Metadata could theoretically be useful, but because of the sheer

volume of images that the organization processes, its use would be impractical. However, House did comment that image technicians regularly adjusted image white point, color balance, and memory colors, all of which require some use of technical metadata. The organization uses Image Performance Assessment tools for testing purposes, and has used these tools occasionally for high profile photo assignments, but their use is not a regular part of the digital imaging process. To maintain a record of change history image technicians perform image adjustments on added adjustment layers in *Photoshop*. The original image file is then saved with these additional layers. House stressed the importance of saving the layers in *Photoshop* as a history of change because the final EPS file is flattened. Alternately, *Photoshop's* history log is sometimes used to keep a record of image adjustments.

House declared that before migrating image files to a new format, the organization would engage in extensive testing to ensure that file integrity was maintained. The New York Times Digital Group has access to the same image archive as the newspaper organization. When repurposing images for the web, the Digital Group does not use any specific technical metadata. Image adjustments are performed separately from the news publication. Additionally, the Times Photo Archive Group repurposes historic images, reprinting them on conventional photographic paper for public sale.

Metadata—Open-Ended Responses

House commented that the most important information the *Times* needed to know about an image was its embedded ICC profile. The organization uses different ICC profiles to convert different types of photographs to CMYK. Interestingly, image technicians work in both RGB and CMYK color spaces when adjusting images. House observed that more experienced workers tended to work in CMYK, while the newer workers were more comfortable working in RGB.

The *Times* keeps a metadata record of intellectual property rights of all its digital images. In addition to rights, the organization keeps a record of the photographer's caption, the printed caption, the story slug or name to match the photo to a story, the editor, the reporter, the photo editor, and the section in which the photo was published. The *New York Times* believes that it is important to establish who worked on an image as well as what work was done to that image. Also, it is important to keep a record of how the original image has been cropped, because there may be valuable information outside the cropped area.

Case Study Five

Introduction

The *Rochester Democrat & Chronicle* is a daily newspaper distributed in the city of Rochester, New York and its surrounding region. The *Democrat & Chronicle* is part of the Gannett Company, which is, by circulation, the largest newspaper group in the United States. According to Gannett corporate figures, the *Democrat & Chronicle* has a

daily circulation of approximately 176,000 subscribers, which increases to 239,000 on Sunday. Dennis Floss, Special Publications Manager, commented that based on *Democrat & Chronicle* figures, approximately 377,000 people read the newspaper on a daily basis. In addition to the daily newspaper, the organization produces several specialty publications that target specific market segments. Furthermore, the organization maintains a website that has over 8 million unique visitors monthly. In 1997, the *Democrat & Chronicle* became the first major metropolitan newspaper in the United States to go to a completely digital imaging workflow.

Digital Imaging Workflow

Floss estimated that staff photographers' capture between 1,800 and 2,700 images a week. He also noted that since the conversion to digital photography, a lower percentage of captured images is entered into the system. However, Floss dismissed the significance of this describing much of the previous film photography as note taking, such as shooting the jersey number of a player. Floss commented, "Our assignment system is electronic. There's an intranet where reporters and editors feed assignments into a database, and from the database, a photo editor will prioritize photo assignments." Photographers frequently edit while in the field, for the most part selecting images to be input into the system. For image processing and image archiving, the *Democrat & Chronicle* uses software designed by Gannett Media Technologies called *Digital Collections*. *Digital Collections* is a browser-based digital asset management system designed to manage newspaper image and text content. Images are entered into *Digital*

Collections as JFIF (JPEG Interchange File Format) files, using the RGB color space inherent with the capture device[4]. The organization is beginning to use Camera Raw files, recognizing the loss image quality that occurs with file compression.

Photographers enter caption information, including address and zip code, using IPTC descriptive metadata fields. Floss stressed that photographers not only enter assigned images, but also enter additional images into the system for unspecified future use. To preserve the original digital captures, they are burned to either CD or DVD. After images are entered into the system, editors place photos on the page as RGB. Subsequently, the digital imaging department converts a copy of the RGB image and makes necessary tonal adjustments. Using the same RGB file, image technicians crop and adjust a second copy for the organization's website. The *Democrat & Chronicle* does not use an ICC color managed workflow. Floss commented that, at this time, with skilled technicians adjusting images, the reproduction quality of the newspaper ranks among the best in the world.

Metadata—Response to Questions

As with all other organizations interviewed, the *Democrat & Chronicle* relies on technical metadata that is classified by NISO as Basic Image Parameters. Floss pointed out that the organization does use Image Creation metadata as a feedback tool to assist photographers. However, he noted that there is no systematic approach to the use of metadata because of the fast pace of the daily news workflow. The *Democrat & Chronicle* does not use Image Creation metadata for image adjustment. At this time,

skilled technicians using mostly visual information and standard industry settings adjust images.

The *Democrat & Chronicle* relies more on Image Performance Assessment tools in its digital imaging workflow than most newspapers. Gray cards, ColorCheckers, as well as standards industry test patterns are used on a regular basis. Additionally, the organization uses Image Performance Assessment tools to adjust the quality of output. Floss declared, "We actually print targets [on a quarterly basis] on press as part of a full production run, and then work on corrections in prepress based on the results." Furthermore, cameras returning from repairs are tested before being put into regular use. Image Performance Assessment results are kept for future reference; however, Floss commented that this has "become less and less critical as cameras have become more uniform and sophisticated." To keep change history, the *Democrat & Chronicle* uses versioning. In addition, the image originally entered into the system is marked as such, archived, and all subsequent work is performed on a copy. At this time, the implications of losing technical metadata when migrating to a new file format have not been discussed.

Metadata—Open-Ended Responses

The *Democrat & Chronicle* has an existing print library of between six and seven hundred thousand photos, and it has been digitizing these prints on an as needed basis. Floss noted that the organization explored digitizing their entire print collection a few years ago, but the cost was prohibitive. It was noted that there was also a film negative

archive that contained possibly ten times as many images that are stored using a chronological filing system.

The organization has discussed the issues that would need to be addressed to print the newspaper at remote locations, such as the customer's desktop printer. A primary concern was preserving the quality of the reproduction when, in a sense, half of the reproduction process is being handed over to the customer. Floss points out "There's metadata that if properly used could allow profiling data to match the output device for what we saw from the sender's end–not necessarily the [reproduction] quality as much as the intention of the image."

The image database of the *Democrat & Chronicle* is not open to the general public, but it can be used by outside researchers for scholarly purposes. Floss commented that the *Democrat & Chronicle* has not has not yet come up with a suitable business model to enable public access. Therefore, the infrastructure, the safeguards, and the copyright protection necessary for complete public access do not exist. However, once a suitable business model is established these other obstacles will be addressed.

Summary of Responses to Questionnaire

The following matrix provides a summary of the responses from the individuals questioned at the five newspaper organizations. Using a range from 1 (unimportant) to 5 (very important). Interviewees were asked to rate the relevance of various technical metadata fields. Two columns of numbers under one newspaper indicated that two individuals in the organization responded to the questionnaire. The last column

represents the average of responses for each category of technical metadata. Survey data from The *Washington Post* was not available at the time of publication. Refer to Appendix C for a complete questionnaire, including discussions of categories and individual fields.

Table 5. Responses to Questionnaire

Basic Image Parameters	**Organization**						
	USAT		Tribune		Post	NY Times	Roch D&C	Avg
Compression Scheme	4	5	4	4		3	5	4.2
Compression Level	3	5	4	4		3	5	4.0
Color Space	5	5	4	4		4	3	4.2
ICC Profile	5	4	3	4		5	3	4.0
YcbCr Coefficients	5	4	3	4		—	3	3.8
File Size	4	5	4	4		4	5	4.3

Image Creation Metadata								
	USAT		Tribune		Post	NY Times	Roch D&C	Avg
Source ID	—	5	3	5		2	5	4
Image Producer	__	4	3	5		2	4	3.6
Host Computer	2	3	3	3		2	3	2.7
Digital Camera manufacturer	3	3	2	4		3	3	3.0
Digital Camera Model	3	3	2	4		2	4	3.0
F Number	__	4	3	4		1	4	3.2
Exposure Time	__	3	3	4		1	4	3.0
Brightness	__	3	3	4		1	4	3.0
Exposure Bias	__	3	3	4		1	4	3.0
Subject Distance	4	3	2	4		1	3	2.8
Metering Mode	4	3	3	4		1	3	3.0
Scene Illuminant	4	4	2	4		3	3	3.3
Color Temperature	4	3	2	4		3	3	3.2
Focal Length	3	3	2	4		1	2	2.5
Flash	3	3	3	4		2	3	3.0
Flash Return	-	3	2	4		1	3	2.6
Back Light	3	3	2	4		1	4	2.8
Exposure Index	3	3	2	4		1	4	2.8
Auto Focus	__	3	2	4		1	2	2.4
Date Time Created	4	5	3	5		4	5	4.3

Image Performance Assessment	**Organization**						
	USAT		Tribune		Post	NY Times	Roch D&C	Avg
Sampling Frequency Unit	3	3	2	3		3	3	2.8
X and Y Sampling Frequency	3	5	2	3		4	3	3.3
Image Width and Length	3	5	2	5		3	4	3.7
Source X and Y Dimension	4	4	2	5		3	4	3.7
Bits Per Sample	5	5	2	5		3	5	4.2
White Point X and Y Value	4	5	2	5		3	4	3.8
Target Type	4	4	2	4		1	3	3.0
Target ID	4	4	2	4		1	3	3.0
Target ID Media	3	4	2	4		1	3	2.8

Change History								
	USAT		Tribune		Post	NY Times	Roch D&C	Avg
Date Time Process	4	4	4	4		3	4	3.8
Source Data	4	4	4	4		2	5	3.8
Processing Agency	4	4	2	4		2	3	3.2
Processing Software	2	3	3	4		3	3	3.0
Software Version	2	4	3	4		3	4	3.3
Processing Actions	3	4	4	4		4	5	4.0
Previous Image Metadata	4	4	4	5		4	5	4.3

Analysis of Matrix

The survey data was averaged using two different methods; first by individual response and second by the average response of individuals in each organization. This was done to determine if organizations where two individuals answered the survey unfairly biased the overall results. Significant differences between the individual and the

organizational method were not found; therefore, averages from the individual method were used.

Based on responses to the questionnaire, most organizations thought that Basic Image Parameter Metadata was important. However, individuals differed in their evaluation how important about specific fields were, such as Compression or ICC Profile. These differences could be attributed to the position in the organization that the respondent held, and how important the particular metadata field was to their function.

Overall, Image Creation Metadata was considered only moderately important. In general, it scored an entire level lower than Basic Image parameters. Notable exceptions were Source ID and Date and Time of Creation, which were rated as important. Interestingly, there was a significant difference between organizations as to the importance of fields containing camera settings. For example, the importance of Focal Length varied between one (unimportant) and four (important).

In general, Image Performance assessment was considered moderately important. Notably, the most importance was assigned to Bits Per Sample. This may be because Bits Per Sample is an important factor in tone and color reproduction. Change History was ranked somewhere between moderately important and important. Surprisingly, the preservation of Previous Image Metadata was rated as very important, with minimal diversity of opinion. The least important area of Change History was considered to be the software used to process the image. Overall, the Change History portion of the questionnaire showed the most agreement among respondents.

Ranking responses by technical category, Basic Image Parameters ranked as the most important category with an average of 4.1. Next, was Change History with an average of 3.6. Image Performance Assessment followed with an average of 3.4, and last was Image Creation Metadata with an average of 3.1.

Endnotes for Chapter 5

[1] A color-managed workflow is where the color rendering of individual input and output devices is characterized according to a device independent standard. Subsequently, the transformation of color data between devices is handled by mathematical transforms or look up tables to ensure consistent color.

[2] Ifra is a worldwide association dedicated to the publishing industry. Its headquarters is located in Darmstadt, Germany.

[3] CCI is a pagination and cross media publishing system specifically designed for the large newspaper market.

[4]JFIF is a subset of JPEG that includes additional headers, so JPEG files can be transferred between a variety of different computer platforms and applications (W3C, 1992).

Chapter 6

Conclusions and Recommendations

Introduction

The purpose of this research was to answer two key questions about the use of technical metadata by the newspaper industry. First, it examined how newspapers use technical metadata in their digital imaging workflow; second, it looked for areas where newspapers could use technical metadata to improve their digital imaging workflow. Chapter Six makes a cross-case examination of the industry's current use of technical metadata. Cross-case examination studies the similarities among organizations as well as the differences between organizations. Next, this chapter examines the use of technical metadata needed to improve digital workflow and the importance of technical metadata in preserving image integrity. In conclusion, this chapter discusses the limitations of the study, and makes recommendations for further research.

Current Use of Technical Metadata

The use of technical metadata under the NISO classification Basic Image Parameters is universal because as NISO aptly states "[this information is] fundamental to the reconstruction of the digital file." It is important to note that, in all instances, this metadata is used through vendor-designed software. However, the use of Image Creation Metadata is not as widespread as Basic Image Parameter Metadata, with two notable

exceptions Source ID and Date and Time of Creation. This is reasonable because these two fields established date and time of capture and assigned each image a unique identifier. Both fields of technical metadata establish traceability to capture device and capture time.

The growing trend of using Camera Raw files has prompted the use of some Image Creation Metadata during adjustment and subsequent conversion of the Raw file to another format. In some companies, Image Creation Metadata is used by photographers to adjust Camera Raw files before they are input into the system; in others editors use it after Camera Raw files are input into the system.

With the exception of *The Rochester Democrat & Chronicle*, Image Performance Assessment Metadata was not used in the organizations studied. Respondents suggested that this is due to the newspaper industry's tight deadlines and a perception of image quality that goes beyond tone and color reproduction. Change History is another classification of metadata that fell victim to the short cycle time in the newspaper industry. For the most part, a history of change was kept at various handoffs in the digital workflow through versioning.

Improving Digital Imaging Workflows

The immediate use of digital images by newspapers requires speed in processing, and an assurance of image integrity. To an extent, the use of technical metadata to streamline the digital workflow is at cross-purposes with newspapers' need to visually examine each image to determine its usefulness. Therefore, it is reasonable to expect that

the most progress in the use of technical metadata to improve the digital imaging workflow will occur through the development of intuitive graphic user interfaces (GUI) that allow for the visual adjustment of images.

The process of repurposing images for other types of media is a growing segment of digital workflow. Repurposing images places different demands on technical metadata, and it is an area where technical metadata could be useful for two reasons. First, technical metadata allows an image that was presented across several types of media to be traced back to its source. Second, because the process of repurposing the same image for different media is repetitive, technical metadata could be used to automate the workflow. The current industry practice of having separate workflows for different types of media is inefficient. Using technical metadata more efficiently creates an opportunity for a single image to have more than one intention.

Another consideration is the use of video clips in newspaper websites. While video is not the same as still digital media, it is reasonable to assume that in the future still images will be extracted from video capture. The extraction of still images from video captures creates a secondary image that will need to be traceable to the source video, which in turn, will need to be traced to the conditions of capture.

To automate the digital workflow, technical metadata must be able to be rendered by machines. Therefore, technical metadata must be included in images using a common machine readable format. The newspaper digital imaging workflow is too fast to trust the inclusion of technical metadata to human beings. Consequently, whenever possible the reliance of user intervention to populate fields of technical metadata should be avoided.

As pointed out by Terrillion from *USA Today*, the inclusion of technical metadata needs to be automated by the software application used to ensure that consistent data is kept on a regular basis.

Another issue brought up by Terrillion was the interoperability of technical metadata with other types of metadata. The use of technical metadata does not occur in a vacuum–it will be used for a purpose, such as accounting, reproduction, repurposing, or archiving. Therefore, technical metadata in digital image files should be designed for use in combination with other types of metadata. Furthermore, to be effective, technical metadata needs to be preserved when several elements on a page are converted to a single document.

There is a trend in large newspaper organizations to create their own image databases for use by other publications within the organization. The need for the inclusion of consistent technical metadata in archives is critical for automatic digital workflows, repurposing of images and the extraction of still images from video capture. Considering the fact that storage will become less expensive and networks will become faster, it would be prudent to include too much technical metadata that could be edited, rather than not to include enough metadata that would require complex workarounds in the future.

Preserving Image Integrity

In all organizations, clear communication of the photographer's intent and a co-operative photo editing process are used to ensure the editorial integrity of the published

image. Editorial integrity is the common denominator in the newspaper industry's definition of quality. It was thought that pleasing images would be the most important quality of a newspaper photograph. However, research pointed out that the definition of image quality in the newspaper industry is far more complex than simple pleasing color. Consequently, at times editorial integrity is achieved by increasing reproduction quality, and at other times it is not.

In all the newspaper organizations interviewed, editors assign staff photographers to a photo shoot. After image capture, photographers use their discretion to determine which images are to be entered into the database. This represents a departure from the traditional film workflow. Previously, when negatives were submitted for processing, there was no pre-editing process; everything that was shot was developed. This change in workflow moves the process of initial editing from the photo editor back to the photographer. Arguably, as Floss of the *Democrat & Chronicle* pointed out, a portion of newspaper film photography was no more than note taking. In spite of this fact, the use of technical metadata between image capture and input into the system would provide traceability to the moment of capture.

The predominant file format used for image capture was JPEG; however, most organizations are moving toward the Camera Raw file format. The major reason for this migration is newspaper organizations' desire to capture as much detail of the event as possible. The adjustment of Camera Raw files through a GUI, and their subsequent conversion to another file format, represents a de facto use of technical metadata to preserve editorial integrity.

With the exception of *The Rochester Democrat & Chronicle,* all organizations use a *MerlinOne* digital management and archiving system. This is logical because newspaper organizations are in the business of providing news, not designing database management systems. Notably, the use of *Merlin* points out the importance of software providers in the digital imaging workflow, and indicates that future development in the use of technical metadata will be through software that they design.

The standard process in all organizations, with the exception of *The New York Times,* does not include ICC color management. In most organizations, the image was placed on the page in RGB color space. Then, the RGB image was replaced by a CMYK image that was adjusted by skilled technicians. It was not made clear exactly why most newspapers did not use ICC color management for editorial photos. Several possible explanations exist, but it is reasonable to assume that since newspaper image quality was a complex matrix of subjective variables other than color reproduction, the benefits of color management were limited. As Floss commented when discussing remote printing, quality is as much preserving the intention of the image as it is preserving the color reproduction of that image. Nevertheless, in most organizations there was some overlap where image technicians also worked on advertising images. Therefore, it would be a mistake for newspaper organizations to assume that ICC color management had no place in the digital workflow.

The needs of the newspaper industry will determine the industry's future digital workflow, and its subsequent use of technical metadata. In all organizations, the newspaper was only one part of a larger media business where news was distributed

across multiple channels. Therefore, any scheme of using technical metadata must take into account the multifaceted nature of these organizations, because, increasingly, the acquisition of news is becoming a separate entity from its channel of distribution.

Limitations of the Study

It was planned to conduct the first interview at *The Rochester Democrat & Chronicle*. Subsequently, interview questions would be adjusted based on those initial results. Unfortunately, due to scheduling difficulties, the *USA Today* interview was conducted first, and *Democrat & Chronicle* interview was conducted last. To compensate for this change, the researcher used the initial set of questions without modification. Therefore, the opportunity to improve interview questions was lost, but the structure that facilitated cross-case comparison was preserved.

Another limitation of this study was that the concept of "technical metadata" was unfamiliar to many respondents. Therefore, the researcher provided NISO technical metadata definitions to increase meaningful responses. Even though these definitions were established in accordance with NISO standards, it is inevitable that the researcher's own opinions had some influence on respondents.

Finally, the reader is cautioned not to over generalize these findings. This study was conducted using data from several of the largest newspaper organizations in the United States. It is not clear whether these findings would apply to smaller newspaper organizations or to the graphic arts industry in general. The answer to these questions needs further study.

Recommendations for Further Research

Since it was found that the use of technical metadata was largely dependent on the applications used by each organization, interviewing software designers would be an excellent subject for further research. A particularly interesting topic would be determining how software manufacturers obtained feedback from newspapers to make decisions regarding software design.

Another area recommended for further study is the use of technical metadata to repurpose images for other media. Newspapers commonly and repeatedly repurpose adjusted images. Therefore, it is recommended that research be conducted in including technical metadata in the automated routines of the digital workflow.

Furthermore, this research only touched on the human element in the use of technical metadata. It would be beneficial to examine human cooperation in, as well as human resistance to incorporating technical metadata in the digital imaging workflow. Also, by implication, this research pointed out the need for technically skilled staff in the newspaper industry. Since it is still common for staff in the newspaper industry to be trained on the job, it can be difficult to introduce new concepts such as technical metadata. Therefore, an area of further research would be to examine the expertise that the incorporation of technical metadata in digital workflow requires, and to make recommendations on how to provide training.

Summary of Implications

As more fields of technical metadata are incorporated into software applications the use of technical metadata in the newspaper digital imaging workflow will most likely increase. A fundamental use of technical metadata will be to automate repetitious image processing tasks, such as repurposing images for web pages. In addition, technical metadata will work in conjunction with other types of metadata to create an enriched interface for the creators and users of digital files. In the foreseeable future, it is likely that metadata displayed by a rendering device will be tailored to the viewer's specific needs.

As newspapers look for alternative sources of revenue, technical metadata will find increased use in digital image archives. There are two avenues in particular where technical metadata will prove useful. First, technical metadata will increase the functionality of image databases, hence increasing their value. Second, technical metadata will help automate the digitization of negatives and prints, reducing archiving costs. Newspaper organizations know that they have a valuable resource in their print archives, but the cost of mining that archive has often proved to be prohibitive.

In spite of automation through technical metadata, the image selection and adjustment process will remain visual. Therefore, the use of technical metadata will increase as software manufacturers develop GUIs that increasingly incorporate technical metadata fields. One possibility is that by using technical metadata, especially those fields classified as image creation metadata, users could compare or even follow an inverse process to create images with different camera settings. However, the last

speculation would create an interesting situation. If technical metadata embedded in an image were used to alter the image, would that image still be the original? To look at it another way, does separating the decisive moment of capture from the capture settings matter? The incorporation of technical metadata into the digital imaging workflow is certain to bring opportunities and challenges to the newspaper industry.

Impact on Newspaper Staff

Technical metadata will make the digital imaging workflow quantifiable; thereby, promoting the use of a system-based workflow. The consequence is that image adjustments will be made earlier in the digital workflow. This phenomenon can already be observed in the adjustment of images by photographers using the Raw format. As a result, the adjustment of images in the production phase of the digital workflow may become obsolete.

Initially, this will have a negative effect on productivity because photographers will spend more time adjusting their creations. However, in the long term this change will be positive, because photographers will insist that their adjustments be preserved throughout the workflow. To meet these demands, production staff will need to adopt a system-based approach to production, which, in turn, should reduce newspaper production costs.

In addition to reducing the need of production staff to adjust images, technical metadata will reduce the time technicians need to search for and repurpose image files, as well as alter the basic skill sets needed for these tasks. In image archives, racks of CDs

and desktop folders will be replaced by searchable databases. Therefore, it is important that newspaper organizations assess, and train employees to keep pace with technical developments.

References

References

Adobe. (2004). White Paper. *A manager's introduction to adobe extensible metadata platform: The adobe xml metadata framework*. Retrieved May 9, 2004 from: http://www.adobe.com/products/xmp/pdfs/whitepaper.pdf

Adobe. (2003). *Working with raw image files in photoshop*. Retrieved November 12, 2004 from www.adobe.com/products/photoshop/ pdfs/ps_cameraraw_userguide.pdf

Adobe Systems. (1992). *TIFF, revision 6.0, final—june 3, 1992*. Retrieved April 1, 2004 from http://partners.adobe.com/asn/tech/tiff/specification.jsp

AdsML. (2004). *AdsML consortium*. Retrieved December 5, 2004 from http://195.52.248.218/WebSite/adsml.nsf/HTML/Index.html

Baca, M. (1998). *Introduction to metadata: Pathways to digital information*. Los Angeles: Getty Information Institute.

Camera Bits. (2004). *Photo mechanic: The essence of editing*. Retrieved November 15, 2004 from http://www.camerabits.com/pages/PM4.html

Castro, E. (2001). *Xml for the world wide web*. Berkeley, CA: Peachpit Press.

Chung, R. (2002, January). *Color management systems*. Lecture presented at the Rochester Institute of Technology, Rochester, NY.

Controlled Vocabulary.com. (n.d.) *The IPTC-NAA standard*. Retrieved March 20, 2004, from http://www.controlledvocabulary.com/imagedatabases/iptc_naa.html

Cost, F. (1997). *Pocket guide to digital printing*. Albany, NY: Delmar Publishers.

Digital Imaging Group. (2000, August 30). DIG35 Specification*: Metadata for digital images, version 1.0*.

Dublin Core Metadata Initiative. (2003, February 6). *Dublin Core Metadata Element Set, Version 1.1*. Retrieved March 28, 2004, from: http://dublincore.org/documents/dces/

Frey, F. (2000). Measuring the quality of digital masters. *Council on Library and Information Resources*. Retrieved August 3, 2004 from http://www.rlg.org/visguides/visguide4html

Hunt, R. W. G., (2004). *The reproduction of colour*. (6th ed.). West Sussex, England: John Wiley & Sons.

Günter, W., & Dale, R. L. (2004, April 20-23). Automatic exposure: Capturing technical metadata for digital still images. *Proceedings of the conference on Archiving: The Society for Imaging Science and Technology*. Springfield, VA.

Idealliance. (2003, July 30). *DISC specifications, and disc metadata specification, version 9*. Retrieved March 20, 2004 from http://www.disc-info.org/specifications/

Idealliance. Digital Image Submission Criteria (DISC). (n.d.). *What is disc?* Retrieved March 28, 2004 from http://www.disc-info.org/about/

Idealliance. *DISC xmp custom panels*. Retrieved March 28, 2004 from http://www.disc-info.org/resources/xmp.asp

Jacquard, J. & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression*. Thousand Oaks, CA: Sage Publications.

Japan Electronics and Information Technology Industries Association. (2002, April). *Exchangeable image file format for digital still cameras: Exif Version 2.2*. Retrieved April 29, 2004 from http://www.exif.org/specifications.html

Jung, K., & Zellmann. T. (2004, April 20-23). JPEG2000/Part6 for scanned documents in archiving applications. *Proceedings of the Conference on Archiving: The Society for Imaging Science and Technology*. Springfield, VA.

Kodak (2002, June). Technical Information Bulletin: *What is an exif jpeg file*? Retrieved July 12, 2004 from http://www.kodak.com/global/en/service/tib/tib4317.shtml

Kodak (2000). *Advantix System: frequently asked questions*. Retrieved August 20, 2004 from http://www.kodak.com/global/en/consumer/aps/faq.shtml

Kodak. (n.d.). Technology Highlights: *Metadata and you*. Retrieved April 18, 2004 from http://www.kodak.com/US/en/corp/researchDevelopment/technologyFeatures/metadata.shtml

Long, B. (2003, March 11). The raw truth: Work with raw digital camera files in photoshop. *Creativepre.com*. Retrieved November 12, 2004 from http://www.creativepro.com/story/review/19008.html

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). London: Sage Publications.

Murphy, E. P. (2004, March 15). *A review of standards defining testing procedures for characterizing the color and spatial quality of digital cameras used to image cultural heritage*. Munsell Color Science Laboratory: Rochester Institute of Technology

National Standards Information Organization (NISO). (2004, March 29). NISO draft standard: *Data dictionary—technical metadata for digital still images, working draft, 1.2*.
Retrieved May 9, 2004 from http://www.w3.org/XML/

Saldago, R. J. (1998, September 26). Abandon film? Tough decisions for photo departments in new era of digital cameras. *Editor & Publisher, 131*. 39 pp. 38-39.

W3 Schools. (n.d.). *Introduction to xml*. Retrieved July 1, 2004 from: http://www.w3schools.com/xml/xml_whatis.asp

W3 Schools. (n. d.) *XSL languages*. Retrieved July 11, 2004 from: http://www.w3schools.com/xsl/xsl_languages.asp

World Wide Web Consortium [W3C] (2004, February 10). RDF primer: W3C recommendation. Retrieved July 12, 2004 from http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

World Wide Web Consortium [W3C]. (2003). *Extensible markup language (xml)*. Retrieved July 16, 2004 from http://www.w3.org/XML/

World Wide Web Consortium [W3C]. (2001). *The semantic web*. Retrieved July 19, 2004 from http://www.w3.org/2001/sw/

World Wide Web Consortium [W3C]. (1992). *JPEG file interchange format, version 1.02*. Retrieved February 27, 2005 from http://www.w3.org/Graphics/JPEG/jfif.txt

Appendices

Appendix A: Interview Questions

1. How many different channels of delivery do you use for news?
   - Please specify

2. How many digital images does your organization take in a year?
   - What percentage of those images is kept, deleted, and printed?

3. Describe your current digital imaging workflow
   - What technical specifications does your organization use for images?

4. How does your organization archive images?
   - What are the uses of the image archive?

5. Does your organization descriptive metadata to categorize images?
   - If so, how is the descriptive metadata kept?

6. Do you think that image creation metadata about *camera settings, and conditions surrounding image captur*e could provide creative and technical feedback for photographers?

7. Do you think that image creation metadata about *camera settings, and conditions surrounding image captur*e could provide useful information for image adjustment.

8. What image performance assessment tools does your organization use?
   - Is image performance assessment information kept for future use?

9. Is image performance assessment data used to control the quality of output?

10. Is a change history of adjustments to digital images kept?

11. When digital images are converted to a new format is there a plan to preserve technical metadata from the previous format?

12. What technical metadata is used when repurposing an image for another media?

13. Are there any areas where you feel technical metadata could be particularly useful, or not useful?

Appendix B: Case Study Questionnaire

Introduction

Metadata is often described as the data about the data. Technical metadata embedded in digital images provides organizations with a powerful tool to manage image data to improve quality, to speed image processing, and to build an organized database for future use. Illustrating the potential of technical metadata in the newspaper industry, Adobe (2004) states "a content composed for newspaper broadsheet will be repurposed for distribution via server to a Web browser on a PC, to a PDA, a cell phone, collected into an archive, and perhaps even pressed into a CD" (p. 15).

The questions in this study were developed using specifications drafted by the National Information Standards Organization (NISO). Currently, NISO describes 111 fields of technical metadata. Below is an abridged list of fields the author considers to be the relevant for the newspaper industry. To download a complete set of NISO standards see references at the end of this document.

Image capture devices automatically record many of the data fields described below; however, some fields, such as processing agency, need to be entered manually.

NISO defines four categories of technical metadata:

Basic image parameters

NISO (2004) defines basic image parameters as data "fundamental to the reconstruction of the digital file as a viewable image on electronically interfaced displays" (p. 13). Basic image parameters include: file type, file size, image resolution, image compression, and embedded ICC profiles.

Image creation metadata

Image creation metadata is data surrounding the image capture and capture system used (NISO, 2004, p. 25). Image creation data recorded at image capture includes: a unique image identifier, the creator of the image, details of the capture device, captures device settings, subject distance, and scene illumination.

Image performance assessment

In describing image performance assessment NISO (2004) claims image performance assessment "has both a present and future context: these elements serve as metrics to access the accuracy of output (today's use), and to access the accuracy

preservation techniques, particularly migration (future use) (p. 35.). High-level spatial metrics enable the accurate recreation of the image when the original source characteristics are not totally available. Furthermore, lower level metrics, such as targets, provide an absolute reference about the image capture.

Change history

Change history is the cumulative record of processing steps that have been made since the capture of an image. NISO (2004) points out "The metadata blocks in change history are used to document the source, systems, and settings used in all subsequent digital-to digital operations" (p. 55).

Below is an abridged listing of NISO technical metadata fields. Please rate the importance of the following technical metadata fields by circling the number that applies:

5 = Very important
4 = Important
3 = Moderately important
2 = Of little importance
1 = Unimportant

Basic Image Parameters

NISO Field	Description	Importance				
Compression Scheme	Compression scheme used to store data	5	4	3	2	1
Compression Level	Level of compression used	5	4	3	2	1
Color Space	Color space of decompressed image	5	4	3	2	1
ICC Profile	Name of embedded ICC profile	5	4	3	2	1
YcbCr Coefficients	Transformation from RGB to YcbCr	5	4	3	2	1
File Size	Number of image bytes	5	4	3	2	1

Are there other (please specify) basic image parameters that are not listed, but if included, would improve your use of digital images?

Image Creation Metadata

NISO Field	Description	Importance				
Source ID	Unique identifier of the source image	5	4	3	2	1
Image Producer	Organizational level producer of image	5	4	3	2	1
Host Computer	Computer and/or operating system at image creation	5	4	3	2	1
Digital Camera manufacturer	Digital Camera manufacturer	5	4	3	2	1
Digital Camera Model	Digital Camera Model	5	4	3	2	1
F Number	F Number used for image capture	5	4	3	2	1
Exposure Time	Exposure time in seconds	5	4	3	2	1
Brightness	Brightness in APEX Values	5	4	3	2	1
Exposure Bias	Exposure bias of capture device in APEX units	5	4	3	2	1
Subject Distance	Distance in meters between subject and camera focal plane	5	4	3	2	1
Metering Mode	Method of meter weighting used	5	4	3	2	1
Scene Illuminant	Light source present at image capture	5	4	3	2	1
Color Temperature	Color temperature of scene illuminant	5	4	3	2	1
Focal Length	Focal length of lens used to capture image	5	4	3	2	1
Flash	Whether a flash was used at image capture	5	4	3	2	1
Flash Return	Camera judgment of flash effectiveness	5	4	3	2	1
Back Light	Amount of back light at time of exposure	5	4	3	2	1
Exposure Index	Camera exposure index setting	5	4	3	2	1
Auto Focus	Use of manual or auto focus	5	4	3	2	1
Date Time Created	Date and time image was created	5	4	3	2	1

Are there any (please specify) image creation metadata that are not listed, but if included, would improve your use of digital images?

Image Performance Assessment

NISO Field	Description	Importance				
Sampling Frequency Unit	Unit of measurement used for sampling frequency	5	4	3	2	1
X and Y Sampling Frequency	Number of pixels per sampling unit in the X and direction	5	4	3	2	1
Image Width and Length	Width and length of image in pixels	5	4	3	2	1
Source X and Y Dimension	Width and length of captured object in selected units	5	4	3	2	1
Bits Per Sample	Example: 8-bit or 16-bit RGB	5	4	3	2	1
White Point X and Y Value	X and Y coordinates of white point on chromaticity diagram	5	4	3	2	1

The following image performance metadata data describes the type of test target used and the method of its use. For example, if a Macbeth Color Checker target were included in the captured scene, it would be considered an internal target. If the same target were captured separately as part of a sequence of images, it would be considered an external target.

Target Type	Internal or external target	5	4	3	2	1
Target ID	Manufacturer of target	5	4	3	2	1
Target ID Media	Media used to create target	5	4	3	2	1

Are there any (please specify) image performance assessment metadata that are not listed, but if included, would improve your use of digital images?

Change History

NISO Field	Description	Importance				
Date Time Process	Date and time of the most recent processing action	5	4	3	2	1
Source Data	Reference or location of source image file from which the processed image was created	5	4	3	2	1
Processing Agency	Organizational level producer of image	5	4	3	2	1
Processing Software	Name of software used to process image	5	4	3	2	1
Software Version	Version of software used to process image	5	4	3	2	1
Processing Actions	An ordinal list of the processing steps performed	5	4	3	2	1
Previous Image Metadata	Documentation of change history and essential metadata to simulate a return to the original image data	5	4	3	2	1

Are there any (please specify) change history metadata that are not listed, but if included, would improve your use of digital images?

References

Adobe. (2004). White Paper. *A manager's introduction to adobe extensible metadata platform: The adobe xml metadata framework*. Retrieved May 9, 2004 from: http://www.adobe.com/products/xmp/pdfs/whitepaper.pdf

National Standards Information Organization (NISO). (2004, March 29). NISO draft standard: *Data dictionary—technical metadata for digital still images, working draft, 1.2*.
Retrieved May 9, 2004 from
http://www.w3.org/XML/

Appendix C: Workflow