

Federal Digitization Moving to Common Guidelines

The U.S. Federal Agencies Digitization Initiative

<http://www.digitizationguidelines.gov/>

IS&T Archiving 2009 Conference

Michael Stelmach

mste@loc.gov

Carl Fleischhauer

cfle@loc.gov

Library of Congress

Washington, DC

The Library of Congress

National Digital Information Infrastructure and Preservation Program (NDIIPP)

www.digitalpreservation.gov

The screenshot shows the homepage of the National Digital Information Infrastructure and Preservation Program (NDIIPP) website. At the top, it features the Library of Congress logo and navigation links for 'ASK A LIBRARIAN', 'DIGITAL COLLECTIONS', and 'LIBRARY CATALOGS'. Below this is a search bar with a 'GO' button. The main content area is titled 'National Digital Information Infrastructure and Preservation Program (NDIIPP)' and includes a graphic of stylized figures holding hands. A sidebar on the left contains a search box and a list of navigation links: 'Digital Preservation Home', 'Importance of Digital Preservation', 'What the Library Is Doing', 'Partners', 'Achievements', 'What You Can Do', 'News & Events', and 'Contact Us'. The main content area is divided into sections: 'Talking about Digital Preservation' with links to 'David Kirsch, University of Maryland', 'Abby Smith, consultant to NDIIPP', and 'American Voices: Digital Preservation'; 'Video Transcripts' with a link to 'Meeting the Challenge'; and 'News & Events' with a link to 'North Carolina States Project Holds Kick-Off Meeting'. A 'Related Resources' section at the bottom left links to 'Standards at the Library of Congress'.

The Federal Agencies Digitization Guidelines Initiative was launched in 2007 under the auspices of the National Digital Information Infrastructure and Preservation Program (NDIIPP) at the Library of Congress.



But it is very much a collaborative effort with participation from a number of federal agencies. I'll just name a few: the National Archives, the Government Printing Office, the National Gallery of Art, the Voice of America, the National Library of Medicine, the Smithsonian Institution, the Department of Defense Imagery Management Operations Center.

FEDERAL AGENCIES
DIGITIZATION GUIDELINES INITIATIVE

SEARCH

Home < Still Image Working Group

→ HOME

→ NEWS & EVENTS

↓ STILL IMAGE WORKING GROUP


- › Participating Organizations
- › Advisory Board
- › Sub-Groups
- › Documents and Guidelines
- › Resources and Industry Standards
- › Provide Comments

→ AUDIO-VISUAL WORKING GROUP

RELATED RESOURCES

- Glossary of Terms
- Sustainable Formats

RSS E-Mail



STILL IMAGE WORKING GROUP

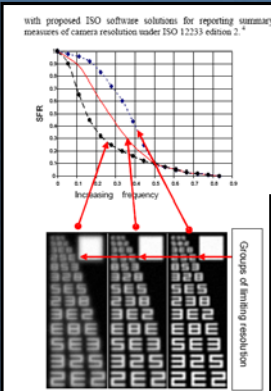
This group is involved in a cooperative effort to develop common digitization guidelines for still image materials (such as textual content, maps, photographic prints and negatives). The expectation is that this work will enhance the exchange of research results and developments, encourage collaborative digitization practices and projects among [federal agencies](#) and institutions and provide the public with a product of uniform quality. It will also serve to set uniform quality and establish a common set of benchmarks for digitization service providers and manufacturers.

The work will focus on guidelines intended for works categorized as historical, cultural and/or archival. In addition to digital imaging and encoding, guidelines will be developed for the metadata that is embedded in digital image files, with a view to increasing the extent to which the files can be "self-describing."

Primary considerations in the development of specific guidelines will be:

- › Defined objectives for the digital object being produced
- › Defined categories and characteristics of content to digitally represented
- › Common image performance measures and methods of validating those measures to defined requirements

Our main emphasis is *digitization*--also known as digital reformatting--the conversion of analog originals into digital form. Michael Stelmach leads the still image group and they are looking at things like books, photos, and maps.



Sharpness vs. Sharpening
 It is often said of digital imaging that *sampling is not resolution*. Image sampling indicates the interval between pixels on a particular plane in the scene (camera), or on the object (scene). Limiting resolution refers to the ability of an imaging component or system to distinguish finely spaced details. Although image sampling (e.g. 500 pps vs. 600 pps scanning) can enable a level of detail in a digital image, it is not the same as, and does not guarantee, the capture of a particular level of limiting resolution. High image sampling is a necessary but insufficient condition for resolving detail.
 Likewise, high (perceived) limiting resolution does not guarantee an overall impression of high sharpness in a displayed image. This was shown recently in Fig. 3 of Ref. 5, part of which is reproduced in Fig. 2. This graph shows the measured spatial frequency response of the two image capture paths from digital

Sharpness Metric
 A typical sharpness metric, CMT acutance \bar{a} is computed as follows. First the system (object-to-display) SFR is measured, then weighted by a one-dimensional CSF

$$\bar{a} = \int_0^{f_{max}} SFR_{system}(f) CSF(f) df \quad (1)$$

This value is then scaled by the corresponding value for an ideal system, where $SFR_{system} = 1$. The visual response value, R , is the ratio

$$R = \frac{\bar{a}}{\int_0^{f_{max}} CSF(f) df} \quad (2)$$

This value is then modified to give the computed sharpness value for the system,

$$CMT = 100 + 66 \log_{10}(R) \quad (3)$$

where R is constrained to the range [0,1].

occurs as the sharpening becomes more aggressive. This SFR bump is a signature behavior of digital sharpening operations. If its maximum amplitude becomes too great, typically greater than 130-150%, over-sharpening artifacts such as halting (Fig. 5) can occur in the image. Exploiting such SFR behavior for measuring sharpness is our proposal. The cited amplitude rules could be one simple technique for measuring sharpness, and in turn monitoring and managing it.



Figure 5: Example of halting effect around characters due to aggressive over-sharpening

Measuring sharpening using SFR.
 Sharpening operations are manifest through the SFR. Among the many reasons for adopting the SFR as a standardized protocol

Measuring and Managing Digital Image Sharpening. By Don Williams and Peter D. Burns, from the IS&T Archiving 2008 conference

Digital still image reformatting is relatively mature and in consequence this group's work--aided by expert consultant Don Williams--moves in ever-more-sophisticated directions.

Advisory Board

DON WILLIAMS



Don Williams has worked as a research imaging scientist at Eastman Kodak in digital imaging since its infancy over 25 years ago. His efforts in the field have concentrated on metrology, image fidelity, quality control, and imaging performance standardization issues especially as they relate to practical workflow adoption. He has published extensively in these areas. He is currently a consultant to the international cultural heritage community and frequently teaches on such matters. Don is the editor for Resolution Measurement.

ROBERT BUCKLEY



Robert Buckley is a Research Fellow with the Xerox Research Center Webster in Webster, NY. He has been with Xerox since 1981, when he joined the Xerox Palo Alto Research Center after receiving a PhD in Electrical Engineering from MIT. He also has an MA in Psychology and Physiology from the University of Oxford, which he attended as a Rhodes Scholar, and a BSc in Electrical Engineering from the University of New Brunswick. During his career at Xerox, he has held several project leadership

STEPHEN ABRAMS



Stephen Abrams is the Senior Manager for Digital Preservation Technology at the California Digital Library (CDL) of the University of California, with responsibility for strategic planning, design, and operation of the CDL's preservation infrastructure. He was the ISO project leader and document editor for the PDF/A standard, ISO 19005-1; the architect and project manager for the JHOVE characterization framework; and the initiator of the Global Digital Format Registry (GDFR) project. He is now leading the multi-institutional JHOVE2 project to develop a next-generation architecture for digital object

Don is one member of the group's advisory panel. The others are Rob Buckley from Xerox and Steve Abrams from the California Digital Library.



Meanwhile, I coordinate the group looking at audio and video. In terms of maturity--or "literacy" to use Don Williams's elegant term--the reformatting of sound recording is catching up with still image digitization, while work with file-based video reformatting and motion picture scanning is just getting rolling in our cultural heritage institutions.

The screenshot displays two overlapping web pages. The top-left page is titled "AUDIO-VISUAL WORKING GROUP Documents and Guidelines" and lists "PLANNED DOCUMENTS" including "Recorded Sound Digitization Overview and Guidelines" and "Video Recordings: Discussion of Digital Target Format Options". The bottom-left page is titled "MEETING NOTES" and contains a PDF link for "Meeting Notes | February 2011 Audio-Visual Digitization". The central page is a navigation menu with links for "HOME", "NEWS & EVENTS", "STILL IMAGE WORKING GROUP" (with sub-links for Participating Organizations, Advisory Board, Sub-Groups, Documents and Guidelines, Resources and Industry Standards, Papers & Presentations, and Provide Comments), and "AUDIO-VISUAL WORKING GROUP". Below this is a "RELATED RESOURCES" section with links for "Glossary of Terms" and "Sustainable Formats". The right page is titled "STILL IMAGE WORKING GROUP Documents and Guidelines" and contains text explaining the group's draft process and a link to "TIFF Image Metadata - Approved".

Both working groups will draft guidelines to support the use of comparable approaches from agency to agency, which will be especially helpful in the relationships with vendors who provide equipment and services.

STILL IMAGE WORKING GROUP
Content Categories and Digitization Objectives

Description of 8 categories and 23 subcategories of printed matter, manuscripts, and pictorial materials. The Working Group's recommended specifications for the digital reformatting of these items (under development) will be defined and evaluated in terms of objectives. Why are the copies being made? What uses will these copies support? Do considerations vary from one category to another? The objectives are presented here as [use cases](#).

CURRENT DRAFT

> [HTML](#) | Content Categories and Digitization Objectives

DOCUMENT BACKGROUND

The Working Group believes that its guidelines should include articulated objectives describing the expected uses of the content. This presentation was drafted by the [Categories and Objectives Subcommittee](#). Each of the 23 content subcategories is reformatting objectives, i.e., the objectives of the organization of the digitizing and, equally important, the objectives will consult or use the images in their work.

CONTENT CATEGORIES

- 1 (T) — Textual and illustrated printed matter (books, journals, manuscripts, some maps). Visual-arts elements of limited significance and generally consisting of printed halftones, line art, explanatory tables and drawings, and the like.
- 2 (PR) — Visual/pictorial items (photographs, prints, some drawings and paintings, some maps). Generally greater visual-art significance than category 1. Two-dimensional, many with continuous tone images (and occasional halftones). Viewed by reflected light.
- 3 (PT) — Photographic negatives and transparencies. Significant visual-arts elements. Viewed by transmitted light.
- 4 (AR) — Special-purpose images (aerial, medical, and scientific images, architectural and engineering line drawings and blueprints). Viewed by reflected light.
- 5 (AT) — Special-purpose images (aerial and medical images). Viewed by transmitted light.
- 6 (3D) — Objects, artifacts, and three-dimensional works of visual art encountered in archives, galleries, and museums (medals and badges, physical evidence from legal archives, some works of art). See also category 7.
- 7 — Specialized imaging of works of art and other objects and artifacts. For future development. To include such examples as two- and three-dimensional works, art in a frame, items with and without gilding, three-dimensional objects in history and science museums, etc.
- 8 — Special-purpose imaging for analysis and research (multispectral, X-ray, other technologies). For future development.

We will define our recommended specifications in terms of objectives. It is not enough to just say “for preservation,” as though that answered the question. The objectives will vary by category of content. We have 8 major categories . . .

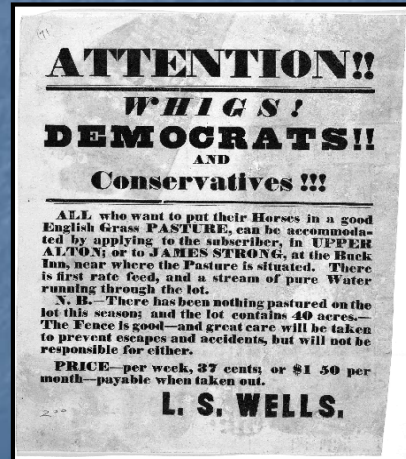
One of the subcategories

T.3. Documents with poor legibility or diffuse characters, e.g., carbon copies, Thermofax/Verifax, etc.; manuscripts or printed/typed pages with handwritten annotations or other markings; items with low inherent contrast, staining, fading, printed halftone illustrations, or included photographs.

. . . that splinter into 23 working subcategories. (We will try to collapse some of these together as we proceed.) Here's one from the still imaging side. We have benefited tremendously from the 2004 imaging guidelines from the National Archives and you may recognize this wording. "Documents with poor legibility or diffuse characters . . . items with low inherent contrast, staining, fading"

One of the subcategories

Valuation: determined by curator or end users to have informational and artifactual value, but not requiring color reproduction.



We added a new distinction: “determined by curator or end users to have informational **and** artifactual value, but **not** requiring color reproduction.” This means that we are not only interested in the words on the page but also in the “page-ness” itself. There are other subcategories for similar content with color, or when an items is of informational interest only.

Selected use case objectives for master images

- Digitizing organization (or successor/receiving agency with an archiving mission) sustains the master (or migrated copies) for the long-term without loss of essential features.

The objectives are stated as use cases and we differentiate between master images and derivative or service images. Here are a couple of for-instances for the master images. One is about longevity--the preservation angle: either this image, or one that we make in a future migration, must carry the underlying content forward over time.

Selected use case objectives for master images

- Digitizing organization uses master to produce derivative images for use cases like these:
 - (1) end-user-access interface
 - (2) other patron uses as listed
 - (3) OCR or other text-creation process
 - (4) document the condition of the original item

A second objective is that the master image is capable of supporting the production of a variety of derivative images. This conception is consistent with Don's semantics yesterday, when he said, "it's an information file, not an image-as-visual." The derivatives (not the master) are where you start thinking about output-referred and see images-as-visuals.

Selected use case objectives for derivative (service) images

- Patron sees inline image or image set in interface. Some view the complete work, a *virtual replica*.
- Patron prints images. Some require print-on-demand copy of complete work, a *physical replica*.
- Patron is confident that the content received is an authentic reproduction, also receives information on restrictions.
- Patron downloads a derivative image and, later, uses embedded metadata to identify content and determined technical provenance.

It's easier to sink your teeth into the use cases for the derivative images. Here are a few for this category, notice that we weave in the need for metadata:

- patron examines images in the online interface
- patron makes a hard copy
- patron is confident that the content is authentic and receives information on rights and restrictions
- patron consults embedded metadata

Selected use case objectives for derivative (service) images

- Publisher uses image to illustrate a book.
- Publisher uses image to illustrate a large poster.
- Exhibit designer uses image for display "mural."
- Broadcaster uses image in high-definition television program, zooming in for Ken Burns effect.

Then there are some objectives that drive quality factors:

- publisher uses image for a high quality book illustration or for a large poster
- exhibit designer uses image for display "mural."
- broadcaster uses image in high-definition television program

Image recommendation in 2004 guidelines from NARA

- 8-bit grayscale mode - adjust scan resolution to produce a QI of 8 for smallest significant character
 - or
- 8-bit grayscale mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger
 - NOTE: Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4,000 lines for 8-bit files

The Working Group has not yet defined its specifications. The 2004 specs from the National Archives get us started with what is needed for derivative images. But a different spin will be in play when we specify the masters . . . the parents of all those derivative and migrated children.

Plan to move from specifications with these factors

- color/monochromatic
- pixel density
- bit depth
- . . . usually output-referred

To specifications with these factors

<u>Tone</u>	<u>Resolution</u>	<u>Color</u>	<u>Uniformity</u>	<u>Noise</u>
<ul style="list-style-type: none"> • Gamma • White Balance 	<ul style="list-style-type: none"> • Spatial Frequency Response (SFR) • Resolution • Sampling Efficiency • Sampling Frequency 	<ul style="list-style-type: none"> • Luminance • ΔE_{2000} • $\Delta E(a^*b^*)_{2000}$ • Channel Mis-registration 	<ul style="list-style-type: none"> • % Lighting Non-uniformity 	<ul style="list-style-type: none"> • Total rms deviation

This has brought us to consider what are the relevant quality factors. We need to move away from things like color vs. monochrome, pixel density, and bit depth, all of which are usually referenced (without saying so) in terms of outputs. Instead we want to be attentive to more appropriate ways to specify tonality, spatial resolution, color, uniformity, and noise.

Bijlage 2. Schematisch overzicht toleranties en criteria Preservation Imaging

Metamorfoze.

	Artwork,	Unique librarian and archival material.	Not Unique librarian and archival material. uniek bibliotheek en archief materiaal	For access only*
Color space	eciRGBv2	eciRGBv2	eciRGBv2	Not specified
Bitdepth	16/8 Bit	8 bit	8 bit	Not specified
Tonalcapture				255
1. pixel value vak A (reflectiewaarde = 0,89)	249 - 238	249 - 238	249 - 238	
2. Highlight gamma	0,8 - 1,08	0,8 - 1,08	0,8 - 1,08	Not specified
3. pixelvalue vak 1 (reflectiewaarde = 0,70)	228 - 218	228 - 218	228 - 218	255
4. D-max	1,95/2,15	1,50	1,50	Not specified
noise	STD < 4	STD < 4	STD < 4	Not specified
Sn/r	> 1	> 1	> 1	Not specified
Uniform illumination	7	7	7	Not specified
• < A-2				
• < A-1	11	11	11	Not specified
• ≥ A-1	13	13	13	Not specified
Colorcast	+3 en -3	+3 en -3	+3 en -3	Not specified
Color accuracy (formula cie 1976)	Gemiddelde ΔE < 4 Maximale ΔE < 12	Gemiddelde ΔE < 6 Maximale ΔE < 15	Niet gespecificeerd	Not specified
hmm	Minimaal 5 hmmm	Minimaal 5	Minimaal 5	Not specified

Working document from the National Library of the Netherlands.

Three columns, three categories. Specifications in the various rows.

Where will the Working Group's specifications for masters end up? Too soon to tell, but here is a page from a very instructive document from the national library of the Netherlands: the columns are categories, the rows are the specs. Hans van Dormolen from the Metamorfoze project discussed this yesterday, very helpful.

Taxonomy of Digital Imaging Performance – Part I
 (Information on evaluation criteria, definitions, related descriptive terms,
 and possible causes of failure can be found in Part II)

Foundation Metrics	Signal								Noise ¹													
Image Science/Engineering Metrics	OEFC (Opto-Electronic Conversion Function)				SFR (Spatial Frequency Response)				Signal-to-Noise Ratio	Radiometric Distortion NPS (Noise Power Spectrum)				Geometric Distortion								
Derivative Metrics	<i>Sensitivity</i>	<i>Tone, Exposure</i>	<i>White Balance / Neutrality</i>	<i>Color Encoding Accuracy</i>	<i>Sampling Rate</i>	<i>Resolution</i>	<i>Sharpening</i>	<i>Acutance</i>	<i>Flare</i>	<i>Depth of Focus</i>	Dynamic Range	<i>Total Noise</i>		<i>Chroma Noise</i>								
												<i>Temporal</i>	<i>Fixed pattern</i>									
												<i>Random (stochastic)</i>	<i>Banding/Streaking (deterministic)</i>	<i>Defects (stochastic)</i>	<i>Non-uniformity (deterministic)</i>	<i>Color Uniformity (deterministic)</i>	<i>Color SFR Uniformity (deterministic)</i>	<i>Regional (deterministic)</i>	<i>Color Misregistration (deterministic)</i>	<i>Aliasing (deterministic)</i>	<i>Spatial SFR Uniformity (deterministic)</i>	<i>Pincushion/Barrel (deterministic)</i>

From this document:

http://www.digitizationguidelines.gov/stillimages/documents/Digital_Imaging_Framework.pdf

Don Williams helped us develop a framework to think through what is needed -- you will recognize it from his talk yesterday. Broadly speaking, it's all about *signal* and *noise*. This document has been posted online and we invite you to send us comments.

Evaluation and Quality Control of Digital Imaging – Part II

- SIGNAL -

SFR – Spatial Frequency Response – (ISO 12233, ISO 16067-1, ISO 16067-2, ISO 15524)
Definition: A spatial frequency descriptor of an imaging system's ability to maintain the relative contrast of input stimuli
Related Concept:
[MTF – Modulation Transfer Function](#)

- SIGNAL -

OECF – Opto-Electronic Conversion Function (ISO 14545)
Definition: Average, large area digital response of an electronic imaging device to light stimuli
Related Concepts:
[TTF – Tonal Transfer Function](#)
[TRC – Tone Reproduction Curve](#)

- NOISE -

- Geometric/Spatial Distortion -

Definition: The deviation of any imaged point from its intended or aim spatial position relative to the input object.

- NOISE -

Radiometric Distortion –

Definition: The deviation of any given spatially imaged point from its aim radiometric value relative to the input object.

Derivative Metrics	Noise Power Spectrum (NPS) Total Noise		Chromatic Noise	
	<i>Definition:</i> A spatial frequency descriptor of the sources of radiometric noise of an imaging component or system		<i>Definition:</i> The cross color channel radiometric deviations relative to an identified aim	
Derivative Metrics	Temporal	Fixed Pattern		Color Uniformity (deterministic)
	Random (stochastic)	Bandwidth/ Streaking (deterministic)	Defects (stochastic)	Color SFR uniformity (deterministic)
Derivative Metrics	<i>Definition:</i> The root mean square deviation (RMS deviation) of two temporal patterns	<i>Definition:</i> One dimensional (1-D) patterns	<i>Definition:</i> point or clusters of defects or poorly connected pixels	<i>Definition:</i> a generally low frequency
Derivative Metrics			Uniformity/ Shading (deterministic)	Color SFR uniformity (deterministic)
		<i>Definition:</i> A difference in large area uniformity/shading between color channels		<i>Definition:</i> The differential spread of light between color channels

More on signal and noise from this document:
http://www.digitizationguidelines.gov/stillimages/documents/Digital_Imaging_Framework.pdf

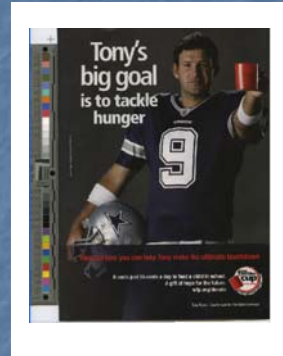
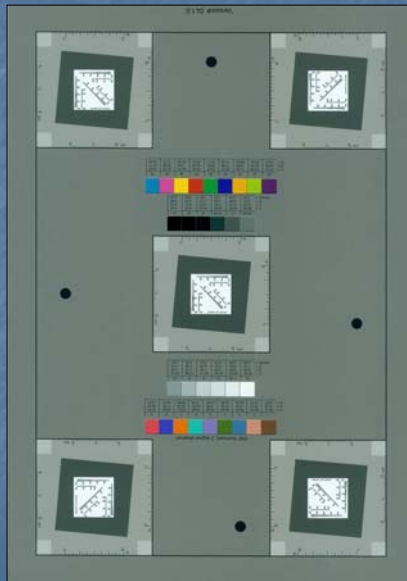
The framework describes four important measurement factors and lists relevant standards documents. For “signal,” we have *spatial frequency response* or SFR and *opto-electronic conversion function* or OECF. For noise, there are several elements under the heading *geometric and spatial distortion* and another set of elements under *radiometric distortion*.

Tools to Support Image Performance Measurement

- Digital Image Conformance Evaluation (DICE) System
 - **Device Target** – Imaging Device Performance
 - **Object Target** – Actual Image Quality
 - **Software** for Evaluation/Validation
 - Based in LabVIEW
 - Data export for use in SQC/SPC

How will you know if your equipment and the work it produces conforms to the recommendations? Don has also been helping us with tools, including a pair of targets and supporting software. (It has been interesting to see the keen interest in targets at this conference. There is increased literacy in the broader community for sure.)

Device and Object Targets



Object target as positioned for use

Thanks to OCLC for help with this part of the effort.

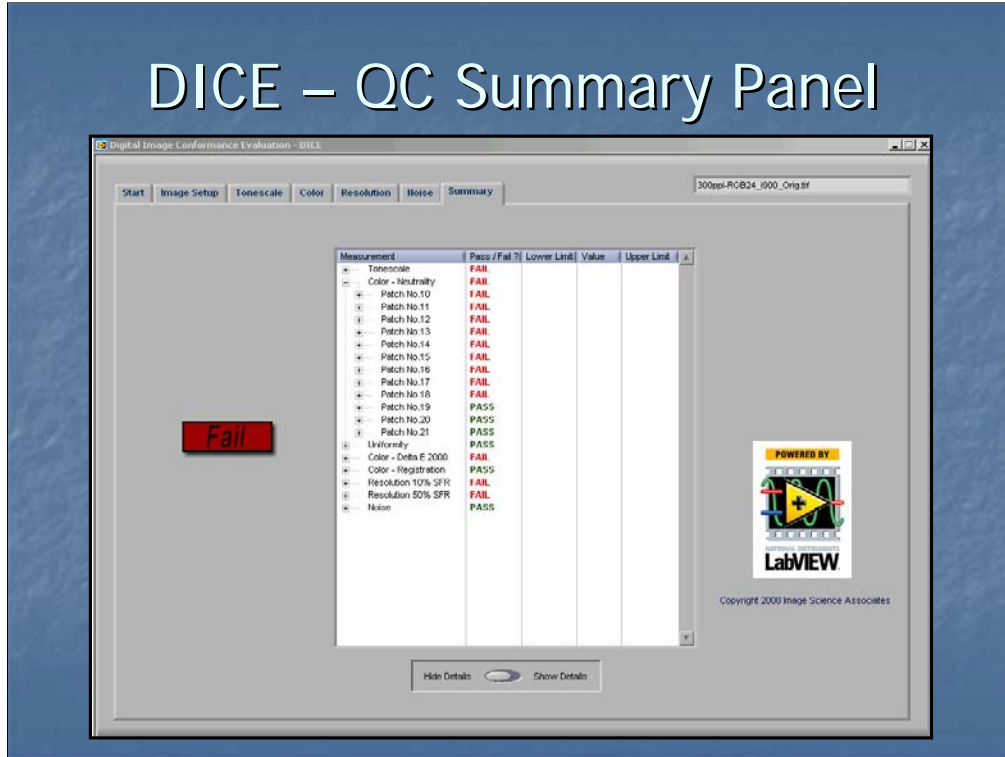
There's a big target for the device and small skinny one, to be imaged alongside the item being reformatted. They are both about ten inches long.

DICE Software – Main Panel

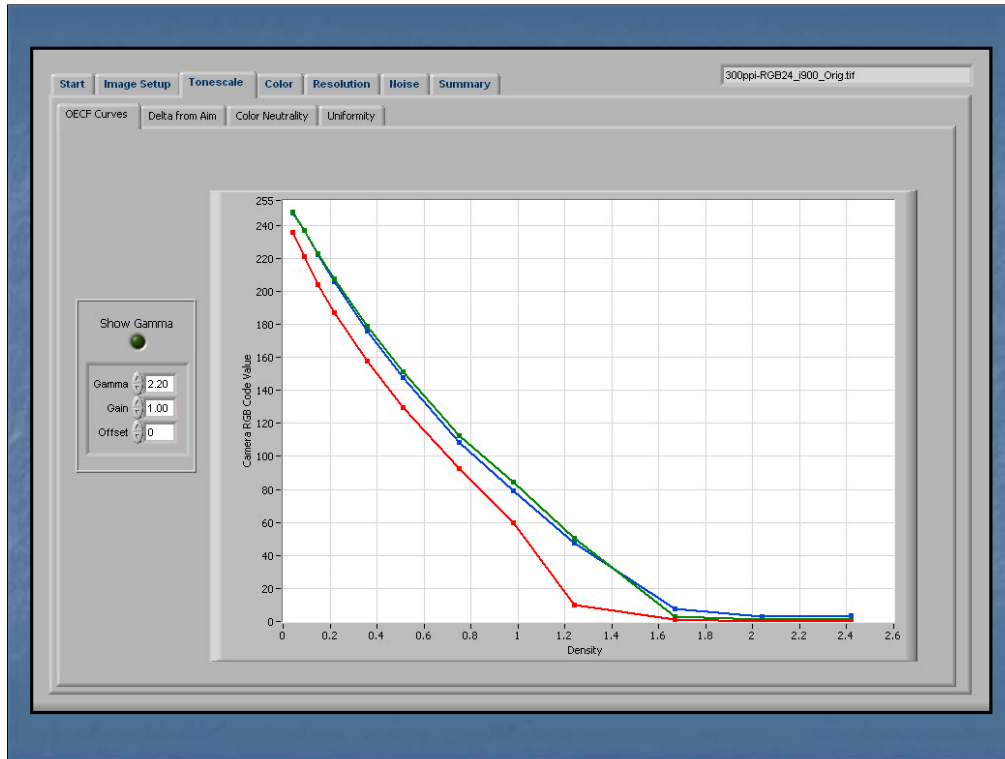


The images of the targets are analyzed by software built on top of LabVIEW, from National Instruments. Our custom package is still in an early beta stage but, when finished, will offer good functionality with an easy-to-use interface. For example, panel number 3 lets you build a job profile, setting the parameters for your current project, what we call “aim points.”

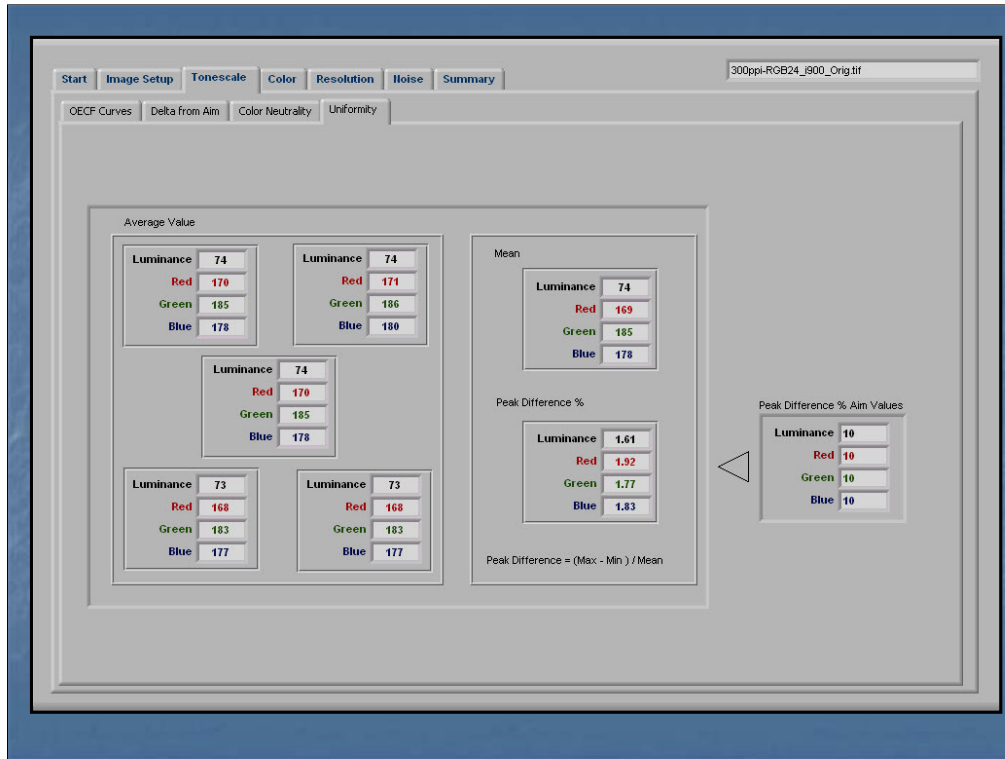
DICE – QC Summary Panel



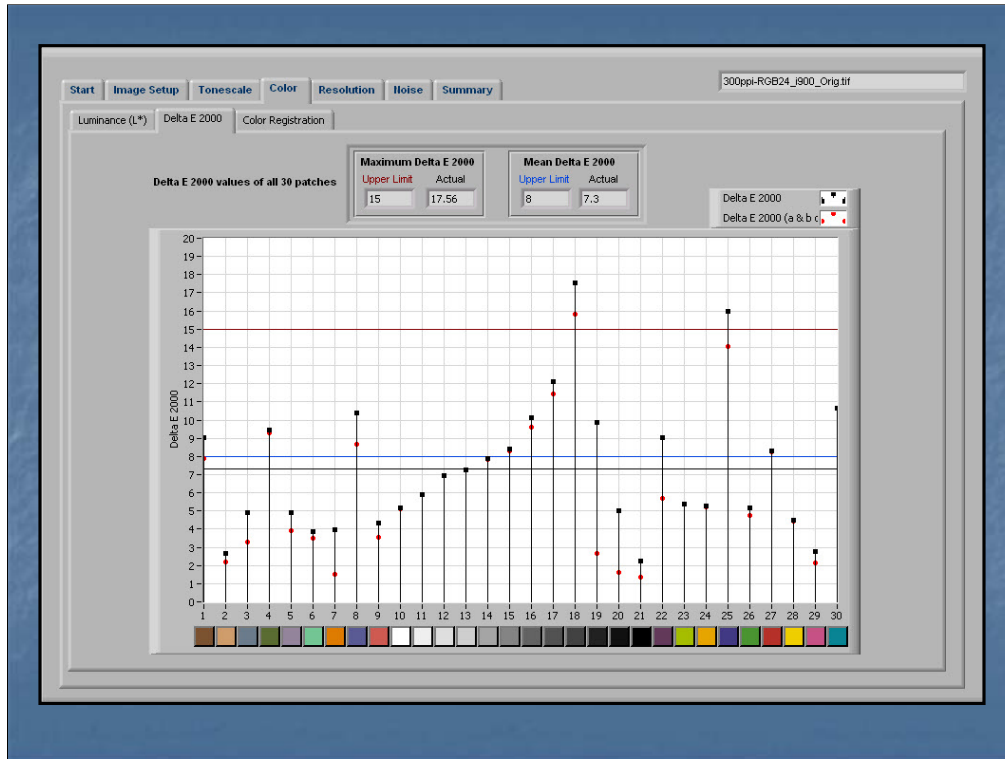
For many workers, the only display of results that they need is the one that tells them that their system passed or--in this case--failed.



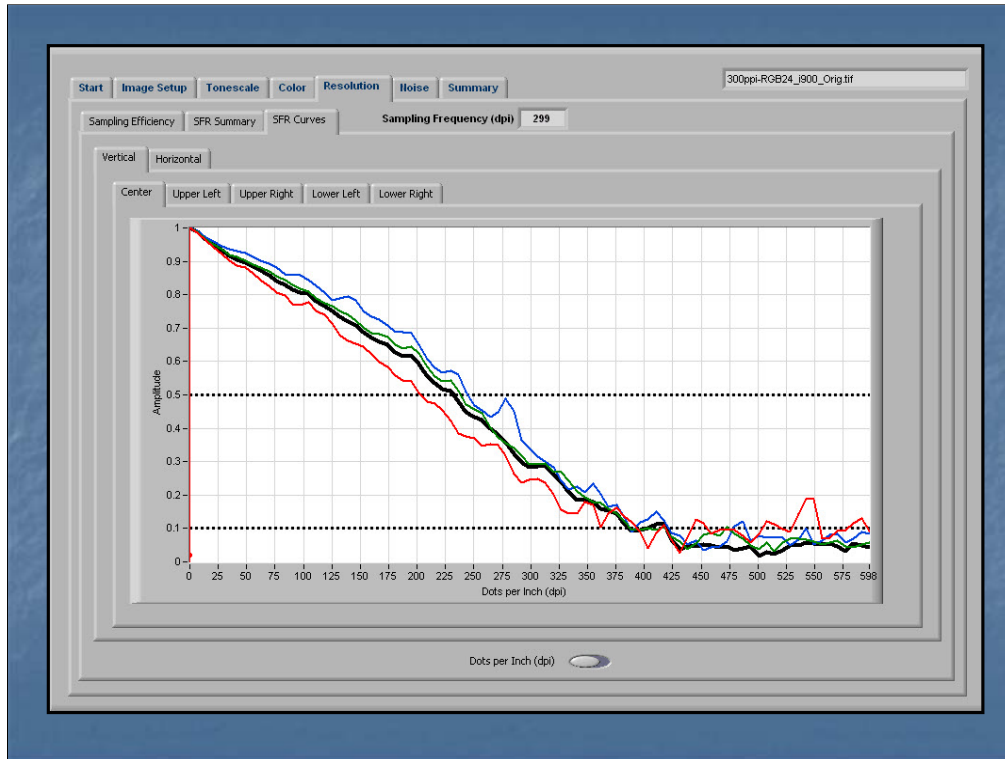
For workers who wish to know more, the software reports details. This display shows the OECF curves, part of what we call “tonescale” measurements.



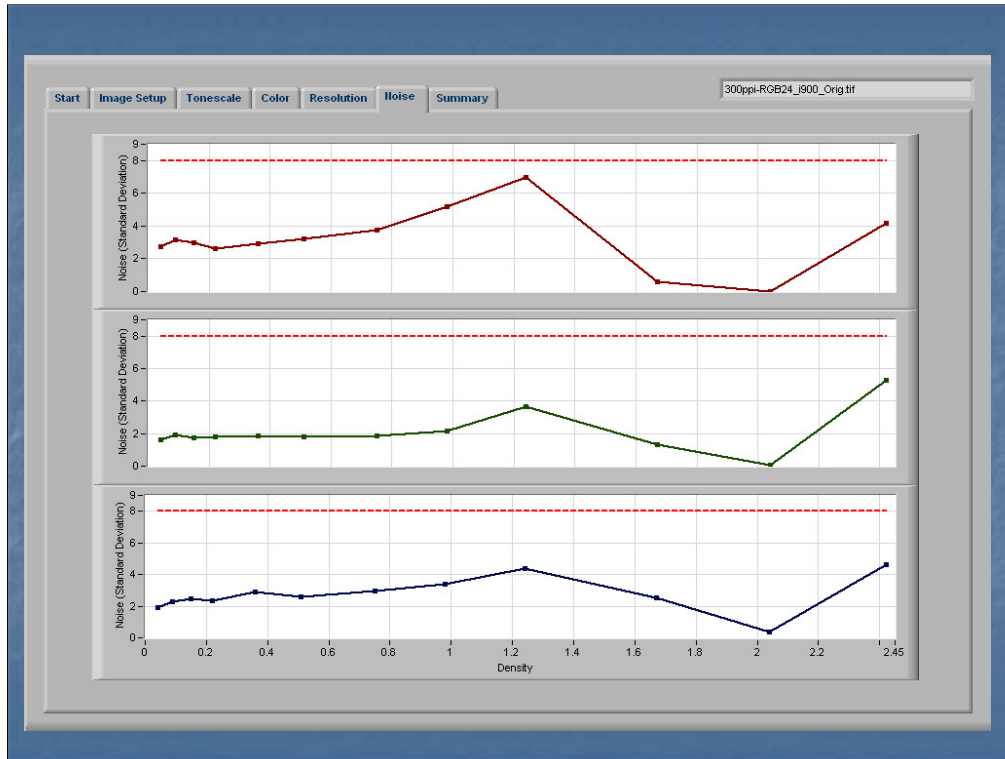
The tabs at the top tell you where you are -- in this case, still in tonescale, this time looking at uniformity.



This display, under color, provides Delta E or color difference information for the calibrated color patches on the target. The horizontal lines show the “aim points” from your profile. In this case, many values lie outside the profile zone; this is one reason that this sample failed.



This one is for resolution, and shows the SFR curve. The upper horizontal line marks the point where the modulation transfer function drops to 50 percent, which we feel is significant for our types of images. Other image-dependent scientific fields use the 10 percent point (lower line) to determine the resolution of their imaging systems.



This display provides information about noise for red, green, and blue channels. This time the values from the target image fall below the level set for this profile--it gets a passing score on this measurement.

Beyond performance measurement

- Embedding metadata
 - TIFF header specification online now
 - Future: exploration of XMP

Guidelines for TIFF Metadata Version 1.0

**Guidelines for TIFF Metadata
Recommended Elements and Format
Version 1.0**

February 10, 2009

Tagged Image File Format (TIFF) is a tag-based file format for raster images. It serves as a wrapper for a variety of encoded bit-plane data. Created in the mid-1980s, TIFF was designed to be cross platform whenever possible, forward compatible. The most recent version is 4.0. Adobe Systems controls the TIFF specification.

Suggested Minimal TIFF Tag Set					
Metadata Set	Tag or Identifier	Name	Description	Sample Values	Note
TIFF tag baseline	256	ImageWidth	The number of pixels per row	1616	Typical scanner size
	257	ImageLength	The number of rows of pixels in the image	4418	Typical scanner size
	258	BitsPerSample	Number of bits per component	8 8 8	Grayscale 24-bit color
	259	Compression	Compression scheme used on image data	1 = Uncompressed 4 = CCITT Group 4	
	262	PhotometricInterpretation	The color space of the image data	0 = WhiteIsZero 1 = BlackIsZero 2 = RGB	Additional color spaces are possible: CMYK (3), YCbCr (6), CIE L*a*b* (6), and others.
	277	SamplesPerPixel	The number of components per pixel	1 3	Grayscale 24-bit RGB color
	282	XResolution	Horizontal pixel count per resolution unit (inches, centimeters)	2400000/10000	240 ppi. Rational data type - 240 ppi/cm is often displayed that way
	283	YResolution	Vertical pixel count per resolution unit (inches, centimeters)	629145400/2097152	300 ppi; see above
	296	ResolutionUnit	Unit of measurement for X and Y Resolution (inches, centimeters)	1 2 3	Some pixels are irregularly shaped Inches Centimeters

Imaging performance is not the only game for the still image Working Group. Another activity concerns the metadata that ought to be embedded in image files. We have published a recommendation for TIFF headers--how to make the most of this rather thin opportunity--even as we begin an effort to explore the ways in which we might make use of XMP in our application.

Beyond performance measurement

- Other “gaps” in prior guidelines to be investigated
 - Image Sharpening
 - Quality Management
 - Image Specification Metric Aims and Limits
 - Foldouts and Inserts in Bound Materials
 - Color Encoding Accuracy
 - Color Space Encoding
 - Selection Criteria for Master Image File Format

There were other “gaps” that we spotted when we reviewed existing guidelines. Here are a few of them, and we will pursue these as possible during the next year or two.

Working draft pertaining to quality assurance and quality control

Work in progress at the National Archives and Records Administration

DRAFT - Quality Control and Quality Assurance of Digitized Files Produced From Partner Projects

I. Introduction

This document defines broad procedures for both Quality Assurance and Quality Control of digitized resources produced primarily from partner/contractor projects.

Quality assurance (QA) is a set of planned and systematic actions necessary to provide confidence that a product or service is meeting specified requirements for quality.¹ A quality assurance program enables the systematic monitoring and evaluation of the various aspects of a project to ensure that pre-def

Quality assurance is generally regarded as a process-driven approach that involves specific steps to help define and achieve acceptable quality levels. Many quality assurance approaches follow the "plan—do—check—act" methodology:

- *Plan*: Establish objectives and processes required to deliver the desired results
- *Do*: Implement the processes
- *Check*: Monitor and evaluate processes by testing the results against predetermined objectives
- *Act*: Apply actions necessary for improvement if the results require any changes

One area for additional development concerns quality assurance and quality review. Here the National Archives has taken the lead. The agency has entered into a number of partnerships with private sector partners. In order to be sure that the digital copies NARA receives from their partners meet appropriate standards, the agency has drafted a quality review specification.

Working draft pertaining to quality assurance and quality control

Work in progress at the National Archives and Records Administration

2.2 Initial Evaluation

The first step in the QA/QC process is to carry out an initial series of benchmarking tests on a sample to test equipment and capture settings, to establish requirements for acceptable quality,

and to establish thresholds for rejection. Supply the partner with sample records to be digitized. If there are multiple record types, supply representatives from each type. Document the results of the initial QA/QC program.

The size of the sample and the number of records to be digitized are determined at this time. Partner(s) agree on a pass/fail basis.

4.2 When during the digitization process is quality control performed?

QA/QC is a phased effort. Although QA/QC procedures should be established prior to the start of a project and followed throughout the entire digitization project, actual inspection and evaluation of digital files often takes place at multiple points in the workflow: after document preparation is completed; during production by the partner after the files have been digitized; and then by NARA after a batch of files has been delivered to NARA. Preliminary quality control of a smaller set of images may take place in a pre-delivery phase, before final transfer and delivery of a batch to NARA.

4.3 What tools are necessary to conduct quality control?

Both pre- and post-delivery quality control should be conducted in a standard viewing environment on a calibrated monitor or approved workstation.

For raster image files, an image viewer application, such as Adobe Photoshop, is necessary for conducting visual evaluation of the images.

Any standard office application (such as Microsoft Excel), or a standard database application, is useful for tracking quality control of digital files.

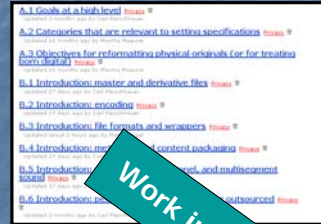
A standard XML editor may be necessary for validation of XML files/schema and for quality control of metadata.

Tools such as JHOVE (JSTOR-Harvard Object Validation Environment) are useful for validation and verification of file formats, embedded metadata, and checksums.

Steve Puglia is leading the NARA team on this topic, and they have shared their draft with the Working Group. This document will provide a starting point for our collective deliberations.

Quick note on audio-visual effort

- Compile guidelines for recorded sound
 - Based on work by IASA and Sound Directions project
 - Parallels to still image effort
 - Embedded metadata (WAVE/BWF, the future)
 - System performance testing, considering use of IASA TC04 pass-fail specifications

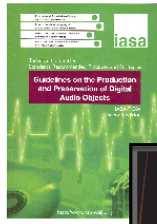


Work in progress



Meanwhile, on the audio-visual side, we have started to compile a guideline--as comprehensive as we can manage-- pertaining to sound recordings.

IASA Technical Committee
 Standards, Recommended Practices and Strategies
 IASA-TC04
 Second Edition

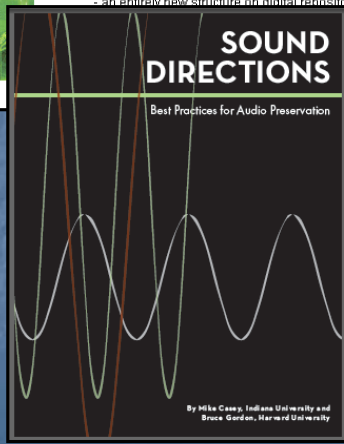


Already an accepted authority in the sound archiving field, this book has been thoroughly revised and updated publication with sub-chapters

The second edition contains:

- guidance in metadata, thoroughly explained with examples
- an entirely new structure on digital repositories which

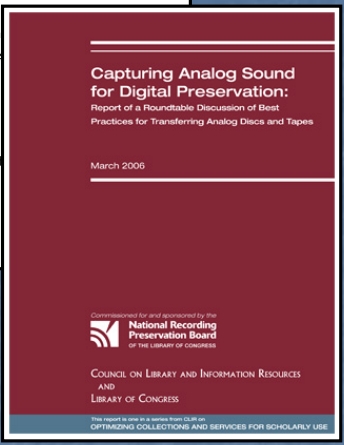
SOUND DIRECTIONS
 Best Practices for Audio Preservation



By Mike Casey, Indiana University and
 Bruce Gordon, Harvard University

Capturing Analog Sound for Digital Preservation:
 Report of a Roundtable Discussion of Best Practices for Transferring Analog Discs and Tapes

March 2006



Commissioned for and sponsored by the
 National Recording Preservation Board
 OF THE LIBRARY OF CONGRESS

COUNCIL ON LIBRARY AND INFORMATION RESOURCES
 AND
 LIBRARY OF CONGRESS

This report is one in a series from CLIR on
 OPTIMIZING COLLECTIONS AND SERVICES FOR SCHOLARLY USE

It will build on the great work produced by IASA--the International Association of Sound and Television Archives--the Sound Directions project from Indiana and Harvard Universities, and the Engineer's Roundtable organized by the National Recording Preservation Board.

Quick note on audio-visual effort

- Video reformatting
 - Plan to document MXF wrapping JPEG 2000 and uncompressed video

MEETING NOTES

Meeting Notes | February 18, 2009
Audio-Visual Digitization Working Group
(PDF, 44KB)

Discussion of film scanning and DPX format; discussion with consultant of planning for three future activities: (i) documentation of specifications for MXF wrapping JPEG 2000 and uncompressed video, (ii) audio-file metadata embedding project, and (iii) testing audio digitization system performance; and discussion of audio production workflow and writing to disk.

This pertains to the exploration of target formats for the reformatting of videotapes. Although no agency has had enough experience to say with confidence, "this format is the one," three federal agencies (LC, NARA, and the Smithsonian) have purchased high efficiency equipment for their current video reformatting efforts. The default output of these devices is an MXF file that wraps (a) picture information compressed with lossless JPEG 2000 and (b) sound information as uncompressed LPCM. But these formats--MXF, JPEG 2000, and LPCM--each allow for some variation in how they are structured. The documentation provided by the manufacturer is helpful on this formatting but it does not provide comprehensive information. Among other things, this lack of complete information inhibits validation.

It is too soon to be comprehensive on the moving image front but--for video--we want to define a profile or application specification for some of the reformatting going on already in three federal agencies. These are efforts that create MXF files, with the picture information in the form of frame images that may be losslessly compressed with JPEG 2000 or left uncompressed.



Regarding motion picture film scanning, we heard from the NASA Johnson Space Center in Houston. As far as we can tell, JSC is doing more of this than any other federal agency. Everyone who scans film seems to use the DPX format, and our Working Group discussion of the matter highlighted some issues. This area could benefit from the development of some best practices.

FEDERAL AGENCIES
DIGITIZATION GUIDELINES INITIATIVE

SEARCH

Home

- HOME
- NEWS & EVENTS
- STILL IMAGE WORKING GROUP
- AUDIO-VISUAL WORKING GROUP

RELATED RESOURCES

- Glossary of Terms
- Sustainable Formats

RSS E-Mail



FEDERAL AGENCIES DIGITIZATION GUIDELINES INITIATIVE

This site is a collaborative effort by federal agencies formed as a group in 2007 to define common guidelines, methods, and practices to digitize historical content in a sustainable manner. Recognizing that the effort would require specialized expertise, two separate working groups were formed with the possibility that more tightly focused groups might be necessary as the work progressed. The [Federal Agencies Still Image Digitization Working Group](#) will concentrate its efforts on image content such as books, manuscripts, maps, and photographic prints and negatives. The [Federal Agencies Audio-Visual Working Group](#) is focusing its work on sound, video, and motion picture film.

To keep up with our progress and submit your suggestions, please visit

<http://www.digitizationguidelines.gov/>

Altogether, there is a lot to do. Michael Stelmach and I hope to hear from you. Thank you for your interest.