

# **FDsys Specification for Converted Content (Version 3.0)**

(June, 2005)

Digitization Specifications and Operating Procedures for  
Archiving Materials: Creation of Preservation Master Files

For the following content types – Textual, Graphic  
Illustrations / Artwork, Originals, and Photographs

**Specifications and metrics for Converted Content – a  
functional solution of the Future Digital System (FDsys)**

***United States Government Printing Office (GPO)***

Final  
June 24, 2005

**FINAL**

**Document Change Control Sheet**

**Document Title:** Digitization Specifications and Operating Procedures

Date	Filename/version #	Author	Revision Description
2/2/2005	<i>DigitizationSpecs-v.1.doc</i>	N. Doyle / R. Selvey	First Draft
2/18/2005	<i>DigitizationSpecs-v1.1.doc</i>	N. Doyle / R. Selvey	Additions, corrections and input from outside sources (LOC, etc.)
2/18/2005	<i>DigitizationSpecs-v2.0.doc</i>	N. Doyle / R. Selvey	Additions, corrections, visuals
3/3/2005	<i>DigitizationSpecs-v2.1.doc</i>	N. Doyle / R. Selvey	Revisions, narrowed down Standards list
4/12/2005	<i>DigitizationSpecs-v2.2.doc</i>	N. Doyle / R. Selvey	Revisions based on workflow
5/10/2005	<i>DigitizationSpecs-v2.3.doc</i>	N. Doyle / R. Selvey	All targets / standards have been established and updated.
5/31/2005	<i>DigitizationSpecs-v2.4.doc</i>	N.Doyle / R. Selvey	Sect III.C – Aimpoints have been revised / updated
6/1/2005	<i>DigitizationSpecs-v2.5.doc</i>	N.Doyle / R. Selvey	Update submission level Metadata
6/24/2005	<i>DigitizationSpecs-v3.0.doc</i>	T. Priebe	Formatted into FDsys template

FDsys Specification for Converted Content (Version 3.0)..... 1

Document Change Control Sheet..... 2

1. Scope..... 4

    1.1 Identification ..... 4

    1.2 Overview ..... 4

2. Referenced Documents..... 5

    2.1 GPO ..... 5

    2.2 Agency ..... 5

    2.3 Industry ..... 5

    2.4 Organizational/Standard ..... 5

3. Current Situation..... 6

    3.1 Background and objectives..... 6

    3.2 Conversion..... 6

        3.2.1 Scanning ..... 6

            3.2.1.1 Current operational situation..... 6

            3.2.1.2 Current Metrics ..... 7

        3.2.2 Inspection..... 11

            3.2.2.1 Current operational situation..... 12

    3.3 Content Management ..... 14

        3.3.1 Image Workflow..... 14

        3.3.2 Asset Management ..... 14

    3.4 Stores ..... 14

4. Desired Situation ..... 14

    4.1 Background Changes ..... 14

        4.1.1 Specific Component..... 14

            4.1.1.1 Objectives ..... 15

            4.1.1.2 Metrics..... 15

            4.1.1.3 Priorities among changes ..... 15

5. Benchmarks..... 15

6. Risks ..... 18

# 1. Scope

What is addressed in this document:

- Scanning and format requirements for text, photographs, and graphic materials
- Digitization Environment
- Digitization Standards
- Required hardware/software configurations
- Quality control

Types of scanning projects will include the following:

- Brittle books (serials and monographs)
- Pamphlets and unbound material
- Archival materials
- Bound materials
- Fold-outs, maps, posters, etc.
- Microform

This specification does not describe how to create a Converted Content Package (CCP). The CCP will be covered in a separate content package specification.

## ***1.1 Identification***

GPO is working with the library community on a national digitization plan for converting the tangible resources held in depository libraries “legacy materials” beginning with the Federalist Papers forward. Digitization of this material will allow wider access to resources and will at the same time provide libraries with the opportunity to reduce the physical volume of their collection. For materials that have previously been digitized by commercial contractors or other organizations, GPO will evaluate these on a case-by-case basis to determine if re-digitization will be required. Factors to be considered will include the availability of the digitized content for free and open access, ability to preserve and create derivatives for content, etc.

The objective is to insure that the digital collection is available, in the public domain, for no-fee permanent public access through the Federal Depository Library Program (FDLP). The digital preservation masters and associated metadata will be preserved, with derivative files made available on GPO Access, and via FDsys once operational.

The end product of the Conversion Process will be a GPO standard Converted Content Package (CCP).

## ***1.2 Overview***

This specification covers all the necessary conversion elements that are required for the creation of a CCP. The components of the conversion solution have been grouped into the following: 1) Conversion Processes; 2) Content Management; 3) Stores.

Converted content is one type of digital content that will be ingested by the Future Digital System. Converted content consists of electronic files created from tangible paper documents, which can be preserved as master files with associated metadata. GPO staff and external service providers “including contractors, library partners, and federal

agencies” will provide converted content to the Future Digital System. The end product of conversion is a Converted Content Package (CCP). The CCP must be produced at a level of quality that is adequate to support preservation as well as future iterations of derivative products.

This document is an outline of our scanning specifications and will continue to evolve and improve as technological advancements occur in the digital imaging industry.

## 2. Referenced Documents

### 2.1 GPO

- Report from the Meeting of Experts on Digital Preservation - March 12, 2004

### 2.2 Agency

Puglia, Steven, Reed, Jeffrey, and Rhodes, Erin. *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files-Raster Images*. College Park, MD: U.S. National Archives and Records Administration (NARA), June 2004. Also available online at

[http://www.archives.gov/research\\_room/arc/arc\\_info/techguide\\_raster\\_june2004.pdf](http://www.archives.gov/research_room/arc/arc_info/techguide_raster_june2004.pdf)

Frey, Franziska S., and James M. Reilly. *Digital Imaging for Photographic Collections Foundations for Technical Standards*. Rochester, NY: Image Permanence Institute, Rochester Institute of Technology, 1999. Also available online at

[http://www.rit.edu/~661www1/sub\\_pages/digibook.pdf](http://www.rit.edu/~661www1/sub_pages/digibook.pdf).

Colorado Digitization Project - *General Guidelines for Scanning*, CDP Scanning Working Group, Spring 1999. <http://www.cdpheritage.org>

Western States Digital Standards Group: *Digital Imaging Working Group - Digital Imaging Best Practices*, Jan 2003.

### 2.3 Industry

- Digital Library Federation's *Benchmark for Faithful Reproductions of Monographs and Serials* (Ver. 1, December 2002)

### 2.4 Organizational/Standard

- The Institute for Museum and Library Services (IMLS) has also published a *Framework of Guidance for Building Good Digital Collections* (2001),

## 3. Current Situation

### 3.1 Background and objectives

The objective of the current situation is to establish a prototype conversion activity to develop workflow processes and metrics to create all conversion elements that are required for the creation of a CCP. GPO will develop the specifications for a FDsys compliant standard CCP as a separate specification.

The As-Is system was designed to test and validate the viability of various technologies and planned processes. DCS is utilizing a pilot operation during its transition period to analyze, develop, and document reporting requirements for the To-Be system. These requirements can then be incorporated into the evaluation criteria for components of the To-Be system and used to evaluate the cost of implementation.

### 3.2 Conversion

Scanning is the only element of the conversion solution that has been benchmarked. Other elements, such as audio and video, need definition.

#### 3.2.1 Scanning

A conversion solution does not currently exist within GPO. Digital Conversion Services (DCS) is currently a prototype operation that is producing scanned images only.

##### 3.2.1.1 Current operational situation

12 Workstations and attached flatbed scanners. Scanning capability is 60 pages per hour per workstation/scanner.

##### Scanning Equipment options:

###### A. Flatbed Scanner

###### 1. Capabilities

- a) Allows the operator to place a single sheet or de-bound materials face down on the scan bed.
- b) Suitable for reflective media (e.g. paper, other substrates).
- c) Suitable for transmissive media such as negatives and film.

###### 2. Limitations

- a) Size limitations based on scanner bed imaging area.
- b) Productivity dependant on operator performance.
- c) Fragile and brittle looseleaf books

###### B. Overhead Scanner/Digital Camera

###### 1. Auto-page turning

- a) *Capabilities*
  - Suitable for bound or non-destructable material.
  - Automated features rely less on speed of the operator.
  - Scans pages while unattended or multi-tasking.

- b) *Limitations*
    - Not suitable for fragile or brittle material.
    - Not suitable for looseleaf or de-bound material.
    - Size limitations based on camera/scanner imaging area.
2. Manual-page turning
    - a) *Capabilities*
      - Suitable for fragile and brittle material.
    - b) *Limitations*
      - Productivity dependant on operator performance.
      - Size limitations based on camera/scanner imaging area.
- C. Auto-document Feed scanner
1. *Capabilities*
    - a) High volume automated processing.
    - b) Suitable for de-bound or destructable material.
  2. *Limitations*
    - a) Scans a limited volume of pages at a time based on the tray size.
    - b) Occasionally introduces distortions due to moving or rotation of pages within the feeder.
    - c) Size limitations based on scanner imaging area.
    - d) Not suitable for rare, valuable, or brittle material.
- D. Film Scanner
- Used for all types of transmissive media (e.g. *microfiche, microform, negatives, and E-6 slides*).
1. *Capabilities*
    - a) Achieves higher resolution necessary for the type/size of media.
    - b) Higher quality and dynamic range.
  2. *Limitations*
    - a) Some film scanners are limited to certain types of media sizes (i.e. 35 mm, medium format, etc), therefore more than one type may be necessary.

### **3.2.1.2 Current Metrics**

Scanning capability for flatbed workflow given existing resources is 60 pages per hour per workstation/scanner.

#### **Environment**

A variety of factors will affect the appearance of images, whether displayed or printed on reflective, transmissive or emissive devices or media. Those factors that can be quantified must be controlled to assure proper representation of an image by its environment.

ISO 3664: Viewing Conditions for Graphic Technology & Photography

*Monitors (refer to NARA Technical Guidelines – pp. 23)*

- The monitor should be set to 24-bits (millions of colors) or greater, and calibrated to a gamma of 1.8 (Mac) or 2.2 (PC).
- Monitor color temperature set to 5000 Kelvin degrees with a desktop background of a neutral gray (avoid images, patterns, and/or strong colors).
- Monitor luminance level must be at least 85 cd/m2 and should be 120 cd/m2 or higher.
- CRT/LCD monitors designed for the graphic arts and multimedia are recommended for a digitization environment.
- Using a target such as the NARA Monitor Adjustment Target or a Kodak Grayscale can be used to adjust the monitor aimpoints of brightness / contrast for calibration (*refer to NARA Technical Guidelines – pp. 24*)

*Room*

- Ambient room lighting should be kept at or below 5000 Kelvin color temperature and should be dispersed/diffused throughout the room, not directly overhead causing glare problems. (*refer to NARA Technical Guidelines – pp. 23*)
- The room should be relatively dust free by use of an air filter and commitment to keeping all scanning systems free of dust and other particles.

**Quantifying Scanner/Digital Camera Performance**

Digitization Standards

Tests should be performed on all image capture equipment prior to purchase and throughout the life cycle of the equipment to ensure quality standards and verification of optimal performance. The following standards should be looked at as benchmarking tools to assess all equipment by either requesting test results from the vendor/manufacturer of imaging equipment or performing an evaluation with the use of a test target for performance metrics. These standards can be purchased from ISO at <http://www.iso.ch> or from IHS Global at <http://global.ihs.com> or other affiliated standards organizations such as ANSI at <http://www.ansi.org/> or AIIM at <http://www.aiim.org>.

Subject	Document Number
<b>Terminology</b>	
Photography -- Electronic still-picture imaging – Terminology	ISO/FDIS 12231.2. July 2004 or 2005
Data Dictionary - Technical Metadata for Digital Still Images (Draft standard for trial use.)	NISO Z39.87-2002 AIIM 20-2002
<b>Opto-Electronic Conversion Function</b>	


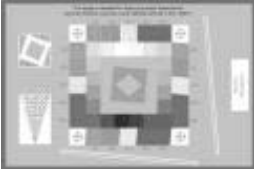
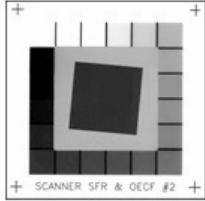

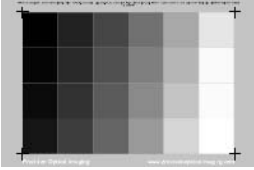



Photography -- Electronic still-picture cameras -- Methods for measuring opto-electronic conversion functions (OECFs)	ISO 14524:1999
<b>Resolution</b>	
Photography -- Electronic still-picture cameras -- Resolution measurements.	ISO 12233:2000
Photography -- Electronic scanners for photographic images -- Spatial resolution measurements -- Part 1: Scanners for reflective media	ISO 16067-1:2003
Photography -- Electronic scanners for photographic images -- Spatial resolution measurements -- Part 2: Film scanners	ISO16067-2 Sept. 2004
Photographic & Electronic Imaging (Resolution definition and application for evaluation of photographic and electronic systems.)	ANSI/AIIM TR26-1993
<b>Noise</b>	
Photography -- Electronic still picture imaging -- Noise measurements	ISO 15739:2003
<b>Dynamic Range</b>	
Photography -- Electronic scanners for photographic images -- Dynamic range measurements	ISO 21550 Sept. 2004
<b>Viewing Conditions</b>	
Viewing Conditions—Graphic technology and photography	ISO 3664:2000
Viewing Conditions—Graphic Technology – Displays for color proofing	ISO 12646
<b>Color</b>	
Photography and graphic technology – Extended color encodings for digital image storage, manipulation and interchange – Part 1: Architecture and requirements	ISO 22028-1:2004
Graphic technology -- Prepress digital data exchange -- Colour targets for input scanner calibration	ISO 12641:1997
<b>Quality Control</b>	
Recommended Practice for Quality Control of Image Scanners. Provides procedures for ongoing quality control of image scanners, including incorporation of targets.	ANSI/AIIM MS44-1988 (R1993)
Sampling Procedures and Tables for Inspection by Attributes. Includes tightened, normal and reduced plans. (American Society for Quality)	ANSI/ASQ Z1.4-2003
Sampling Procedures and Tables for Inspection by Variables for Percent Nonconforming (American Society for Quality)	ANSI/ASQ Z1.9-2003
Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) & Micrographics Systems. Provides guidance in selecting a sampling procedure	ANSI/AIIM TR34-1996

### Test Targets

Before the purchasing of new digitization equipment and after the purchase, an initial performance capability evaluation should be conducted with each digitization device. This may involve using test targets to make benchmark assessments in image quality to predict the integrity of such devices and how effective they will be. Tests are also performed to optimize the performance of an image capture device based on operational settings. These test results should be cumulated into a database to track the performance and/or any variability.

### Targets used for Benchmark Testing Digital Image Capture Devices

Digital Reproduction Elements	Purpose
<p><b>ISO 12233:2000</b>  <b>ISO Resolution Chart for Electronic Still Cameras</b></p> 	<p>Targets: ISO 12233 Resolution Chart  <i>(1X- 35.6cm x 20cm- Chrome on Photopaper)</i></p> <p><a href="#">Link to Purchase</a></p> <p>Designed to check resolution and spatial frequency response of electronic still imaging cameras, this chart comes in a variety of sizes and has testing software available upon request.</p>
<p><b>ISO 16067-1: 2003</b>  <b>ISO 16067-2: Sept. 2004</b>  <b>ISO 14524:1999</b></p> <p><b>ISO Scanner Test Chart for Reflective/Transmissive Scanners</b></p>  <p><b>Slant Edge Target</b></p> 	<p>Targets: QA-61  <a href="#">Link to Purchase</a></p> <p>Determines reflective light resolution and imaging characteristics of digital scanning systems.</p> <p>Targets: QA-62  <a href="#">Link to Purchase</a></p> <p>Designed for evaluation of the slant edge target and used for MTF analysis of the digital scanning system's spatial frequency response (true resolution).</p>
<p><b>Grayscale (Q-13)</b></p> 	<p>Target: Q-13 (small) <i>(comes with Kodak Color Control Patches)</i></p> <p><a href="#">Link to Purchase</a></p> <p>This target can be used to verify if the tonal curves are within a defined range of densities for highlight, midpoint, and shadow. The additional color patches can be used to monitor the calibration (<math>\Delta E</math>) of the imaging capture device and it applies to both monochrome and color electronic still picture cameras and digital scanners.</p>
<p><b>ISO 21550</b>  <b>Dynamic Range Chart</b></p> 	<p>Target:  <a href="#">Link to Purchase</a></p> <p>This International Standard defines methods for measuring the ability of scanning devices to capture tones focusing on the dark areas of the source image. This standard uses digital analysis techniques for measuring Dynamic Range for film and reflective media.</p>
<p><b>ISO 12641:1997</b>  <b>Color Reproduction Target for Calibration</b></p> 	<p>Target: ANSI IT8.7/1-1993 (Kodak Q-60E3)  <a href="#">Link to Purchase</a></p> <p>Transmissive Target for scanner calibration</p> <p>Target: ANSI IT8.7/2-1993 (Kodak Q-60R1)  <a href="#">Link to Purchase</a></p> <p>Reflection Target for scanner calibration</p>

### 3.2.2 Inspection

In the prototype environment, all scanned images are manually inspected.



Document Inspection prior to scanning

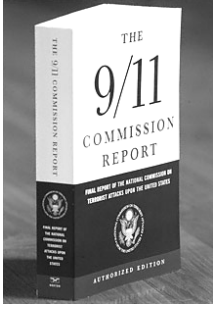

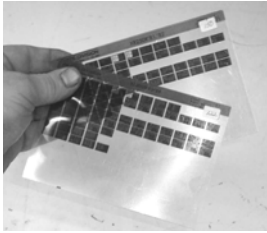
- Determine that all pages are in each publication.
- Determine if there is any damage to publications:
- Torn pages
- Damaged spine
- Stains
- Smudges
- Wrinkles

#### 3.2.2.1 Current operational situation

2 workstations are dedicated to inspection. Inspection is a manual examination of the page as compared to the image.

#### Document Characterization

Categories of Material	Handling	Types of Scanners
<p><b>Type A:</b> <b>Rare, valuable, &amp; brittle</b></p> 	<ul style="list-style-type: none"> <li>• Must be specially handled with white, static-free gloves and treated with care.</li> <li>• Pages turned carefully and book must not be mishandled or dropped.</li> <li>• All areas kept free of extraneous paper dust and dirt through careful measures such as, compressed air or by lightly dusting over the imageable surface.</li> <li>• Some documents may require a translucent protective sleeve prior to digitization.</li> </ul>	<ul style="list-style-type: none"> <li>• Overhead Scanner/Digital Camera – Manual-Page Turning <b>ONLY</b></li> <li>• Flatbed Scanner</li> </ul>
<p><b>Type B:</b> <b>Pamphlets, unbound</b></p> 	<ul style="list-style-type: none"> <li>• Can be separated and run through an automated feed process.</li> <li>• Can be unfolded and placed flat on an imageable surface.</li> <li>• Some may require removal of binding materials (<i>ie. staples, stitches, spiral, comb-binding, tape, etc.</i>)</li> </ul>	<ul style="list-style-type: none"> <li>• Auto-document Feed scanner</li> <li>• Flatbed Scanner</li> </ul>

<p><b>Type C: Bound</b></p> 	<ul style="list-style-type: none"> <li>• Publications scanned while intact and in its original bound form.</li> <li>• Can be opened and placed flat on an imageable surface.</li> </ul>	<ul style="list-style-type: none"> <li>• Overhead Scanner/Digital Camera – Auto/Manual-Page Turning</li> </ul>
<p><b>Type D: Fold-outs, maps, posters</b></p> 	<ul style="list-style-type: none"> <li>• Can be separated and run through an automated feed process.</li> <li>• Can be unfolded and placed flat on an imageable surface.</li> <li>• Some are larger formats and may require a larger scanner/camera imaging device to capture the whole area.</li> </ul>	<ul style="list-style-type: none"> <li>• Flatbed Scanner</li> <li>• Wide Format Cameras/Scanners</li> </ul>
<p><b>Type E: Microform</b></p> 	<ul style="list-style-type: none"> <li>• Many different formats/sizes that may require specific equipment or handling, therefore more than one type of scanner may be necessary.</li> </ul>	<ul style="list-style-type: none"> <li>• Film scanner (<i>various types</i>)</li> <li>• Flatbed Scanner</li> </ul>

**Text Quality (OCR processing)**

1. Determine level of text quality for OCR using a visual scale
  - All typefaces in the publication over 6pt
  - All or some typefaces in the publication under 6pt

**Image Capture Classification**

2. Determine the type of image capture mode performed on each page
  - RGB** (Color halftones, solid images, photographs, charts, or any type of continuous-tone image)
  - Grayscale** (Non-color halftones, solid images, photographs, charts, or any other type of continuous-tone image)
  - Bitonal** (Black and white only – text matter or line-art matter)

### **3.3 Content Management**

#### **3.3.1 Image Workflow**

Currently DCS utilizes a manual process for file workflow tracking and management.

The product set selected by DCS will support document/data capture and production/ad-hoc scanning in a single application. The application will also have a strong Application Programming Interface (API) to expand functionality when needed within the functionality of the COTS product selected. Most structured and unstructured documents can be scanned in batches, and the system should have the capability to automatically recognize each document in a batch and process them based on characteristics that have been predefined. The batch definition process should be full-featured yet simple and quick to use. The product's workflow should be integrated and manage documents allowing a high level of control over how the diverse types of documents that GPO will manage are processed.

The selected product set should combine both document and data capture and allow remote Internet-based capture for future use. Capture stations should be easy to set up at GPO's HQ site and at possible remote sites—across geographic regions or in the same building—and should be able to synchronize with a central capture site via the Internet. It is important that the product selected have an open architecture that makes it easy to extend the basic application to handle complex, high-volume document processing. The product should also be able to predefine “batch definitions or classes” to allow all classes and types of documents to be captured.

#### **3.3.2 Asset Management**

Currently a manual process, files are located based on structured DCS workflow process and file storage scheme.

### **3.4 Stores**

Storage of scanned images are on a network server, with standard IT back-up processing in place.

## **4. Desired Situation**

### **4.1 Background Changes**

Create a scanning environment that incorporates automated workflow software, and combinations of scanning equipment and efficient software to support each area within the workflow.

#### **4.1.1 Specific Component**

A Scanning module should be available to create batches, scan and import documents, and edit the contents of batches. After the batches are created, they should be able to be entered into temporary storage in the system, making them available for processing by subsequent modules.

- **Batch creation:** The operator creates the batch by selecting the type of batch to create (the batch class) and then scanning or importing documents and pages. The document images are stored in a temporary folder for further processing by the system.
- **Batch editing:** Once the batch is created, the operator can visually check documents or pages, and edit them as necessary. Editing functions include replacing, reordering, or rejecting documents and pages. Entire documents or individual pages can be rotated and saved in the rotated state.

#### 4.1.1.1 Objectives

To design a system that constructs as many “mini” conversion pipelines that can stand on their own should a failure occur. Each of these mini pipelines or “clusters” contain workflow, scanning, recognition, key-from image, key from paper, QA, storage functionality and the people to staff its stations. All of the clusters are then managed by a site-level workflow manager which normally manages workflow for all of the clusters, provides administrative functions and communicates with sites and services outside of the confines of the current site.

The system will be broken down into as many “independent clusters” as required to help guarantee reliability. Workflow and administrative functions at the site level will also be organized in a way to make sure that Backups and Administrative tasks are built to make a cluster as independent as possible.

#### 4.1.1.2 Metrics

Metrics of workflow will follow previously mentioned ANSI and ISO standards.

#### 4.1.1.3 Priorities among changes

- 1) Workflow Software
- 2) Batch Processing for Digitization of Documents
- 3) Quality Control Process
- 4) Process for Metadata Capture

## 5. Benchmarks

**Image Capture Benchmarks for Preservation Masters** (refer to NARA Technical Guidelines – pp. 32-36)

*Scanner Setup (refer to DLF – pp. 3, NARA-pp.52)*

Image Types	Bit Depth	Color Mode	Resolution (ppi/spi)	Scale	File Format
<b>Reflective</b>					
<i>B&amp;W Text Only</i>	1-bit	B&W (bitonal)	600 ppi/spi	100% (1:1)	TIFF (lossless or uncompressed)
<i>B&amp; W Text with Illustrations (charts, artwork, graphs, photos)</i>	8-bit	Grayscale	400 ppi/spi *		
<i>Color Photos &amp; Illustrations with Text</i>	24-bit	RGB	400 ppi/spi *		

<b>Transmissive</b>					
16mm	36-48 / 16 bit	Color / Grayscale	5000 ppi/spi	1600% (16:1)	TIFF (lossless or uncompressed)
35mm	36-48 / 16 bit	Color / Grayscale	3400 ppi/spi	850% (8.5:1)	
2-1/4"	36-48 / 16 bit	Color / Grayscale	1800 ppi/spi	450% (4.5:1)	
4" x 5"	24-48 / 8-16 bit	Color / Grayscale	800 ppi/spi	200% (2:1)	
8" x 10" +	24-48 / 8-16 bit	Color / Grayscale	400 ppi/spi	100% (1:1)	

\* Scanning resolutions for images over 11 x 16" (300 ppi for 8-bit grayscale and 300 ppi for 24-bit RGB color)

1. Originals will be backed with bright white opaque paper for flatbed scanning.
2. **Scan Kodak Grayscale Target (Q-13 or Q-14)**, or an equivalent 14-step or 20-step grayscale, only on publications required to preserve color/grayscale data and to further evaluate of the tonal/dynamic range of the scanning device output.
3. **Choose best defined presets to digitally capture type of publication – Based on all these factors:**
  - a) *Color Mode* – to best define the color of the original publication format.
  - b) *Scaling* – to best define the digital capturing parameters according to *III.A Scanner Setup specifications*.
  - c) *Size/Crop* – assuring that an area of at least 1/4" outside of the parameters of the open page(s) is captured.
  - d) *Resolution* – using the correct amount of this is dependant on the type of media as well as the content itself according to *III.A Scanner Setup specifications*. (ie. *transmissive vs. reflective, color vs. grayscale vs. bitonal*)
  - e) *Descreen* – to remove any printed halftones that cause the obtrusive moiré patterns when digitally capturing from printed material such as newsprint or magazine-type paper.
  - f) *Paper/Print Mode* – to determine the optimal settings for the scanner/camera to capture the best rendering of the original (ie. *Some scanner API's have substrate mode [magazine/coated, newsprint, uncoated, photograph] to choose from for the purposes of descreening or other capture features*)
  - g) For significant embossed seals / images, the flatbed scanner must be set use One Directional Light
  - h) *Tonal Adjustments* – scanner hardware and software must be equipped and capable of capturing correct highlights/shadows without losing detail. Also, the software should use tools with more controls (Levels and curves) along with numeric feedback.
  - i) *Color management* could be involved in any settings using proper calibration software for both monitors and image capture devices (Cameras and scanners).

**NOTE:** Presets will be programmed for each scanner based on all these definitions.



*Curvature Reduction*

If available in the API (Application Programming Interface) of the scanning software, applying an in-process setup to reduce the curvature or rotation of pages during the scanning phase may be necessary.

*Aimpoints for Grayscale Target (Tone Compression)*

On the preservation master file, the original scan contains a grayscale target. Tone compression is a technique to make the digital reproduction to look like the original in terms of the exact tonal range.

**NOTE:** This theory should not be applied in all cases, due to each publication’s variation in quality attributes due to aging, or the process used in the creation of the publication.

**Scanning Aimpoints for Grayscale Target (Q-13) using 8-bit Grayscale Mode** (refer to NARA Technical Guidelines – pp. 35)

		Neutralized White Point	Neutralized MidPoint	Neutralized BlackPoint
Step or Density	Kodak Q-13/14	A	M	19
	Visual Density	0.05 – 0.10	0.75 – 0.85	1.95 – 2.05
Aimpoint	RGB Level	242-242-242	142-142-142	22-22-22
	% Black	4%	52%	92%
Acceptable Range	RGB Level	236 – 248	136-148	16-28
	% Black	2 – 6%	50 – 54%	90 – 94%

**Scanning Aimpoints for Grayscale Target (Q-13) using 24-bit Color Mode** (refer to NARA Technical Guidelines – pp. 35)

		Neutralized White Point	Neutralized MidPoint	Neutralized BlackPoint
Step or Density	Kodak Q-13/14	A	M	19
	Visual Density	0.05 – 0.10	0.75 – 0.85	1.95 – 2.05
Aimpoint	RGB Level	242-242-242	122-122-122	40-40-40
	% Black	4%	60%	90%
Acceptable Range	RGB Level	236 – 248	116-128	34-46
	% Black	2 – 6%	58 – 62%	88 – 92%

**Aimpoint Variability** (refer to NARA Technical Guidelines – pp. 34)

For the three aimpoint values described above, none should exceed a variability of ± 6 RGB increments per each individual channel: Red, Green, and Blue. You can verify this by using an image sampler in the scanner software tools or an eyedropper tool from image processing software (such as Adobe Photoshop or equivalent) and set to measure an average of either 3 x 3 or 5 x 5 pixels to sample on the grayscale.

**Note:** never use a point sample or single pixel sample to base your measurement on.

*Verification and Save*

**Check the results of the scan** – In the case that a scan preview varies from the expected results – the scanner settings may need adjustment. If an unknown discrepancy appears on

the scan such as dust, scratch, or any other sort of mark on the digital preview, examine the platen glass to remove any marks, smudges, or dust off the image area and rescan.

**Minimum (submission) level Metadata** - Each publication scanned and digitized, must have a minimal level of metadata associated with each TIFF file for preservation purposes. The data elements will consist of bibliographic, technical, and administrative information necessary to track, manage, and preserve the associated files with each title for the future content management system. The TIFF data elements and values (e.g. presented in XML as fields with values associated with file header tags), represent metadata used to render and manage image data.

**GPO submission level metadata will capture:**

- (1) Identity
  - (a) Title or caption
  - (b) Unique Identifier (persistent locators, filenames, ISNs, etc)
- (2) Responsibility
  - (a) Author / Creator
  - (b) Publisher / Authority
  - (c) Rights Owner \*
- (3) Version / Fixity\*
  - (a) Version information
  - (b) Relationship to other version or manifestations
- (4) Representation / Technical / Structure\*
  - (a) Must incorporate NISO Z 39.87-2002 technical metadata for digital still images
  - (b) Structure Information

\*If readily available

**2. File Naming Convention** –The system identifier requires machine or human indexing for corresponding files that relate to each document. Through a standard naming convention, the process of ingest, storage, search, and retrieval of documents is simplified.

**3. Converted Content Package (CCP)** – The images may be in RGB, Grayscale, or Bitonal mode and should have a unique identifier and metadata associated with each file.

## 6. Risks

- Not incorporating automated workflow software will constrain throughput.
- Not upgrading scanning equipment capability will constrain document scanning options.
- Not automating Quality Control process will increase personnel required, and constrain throughput.