

Digitizing Books: Practices and Guidelines

**Federal Depository Libraries and GPO Conference
December 12, 2013**

Carl Fleischhauer
cfle@loc.gov
Library of Congress
The U.S. Federal Agencies Digitization Guidelines Initiative (FADGI)
Washington, DC

What is a digitized book?

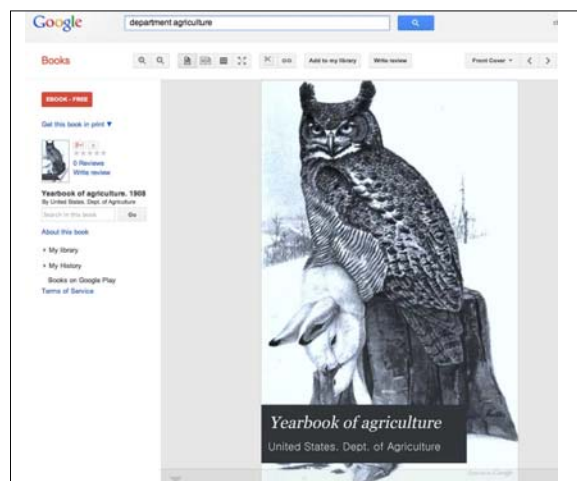
- What is a digitized book?
 - Different from a newly made digital eBook.
- What are its parts?
 - Scanned raster images of pages are central.
- What guidelines or standards come into play?

What is a digitized book?

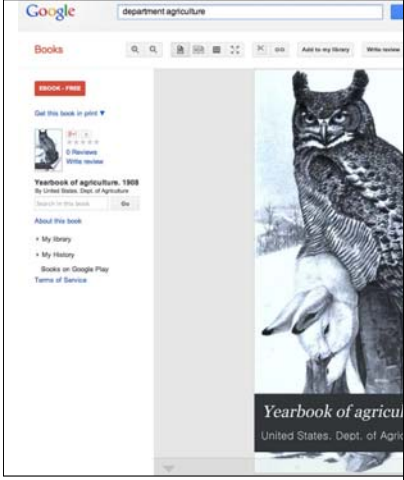
- This talk is about *preservation*, and concerns master files for long-term management
- Nevertheless, online access (also over the long-term) is the primary driver.
- *Let's work back from some examples today. What does a reformatted, online book consist of?*

3

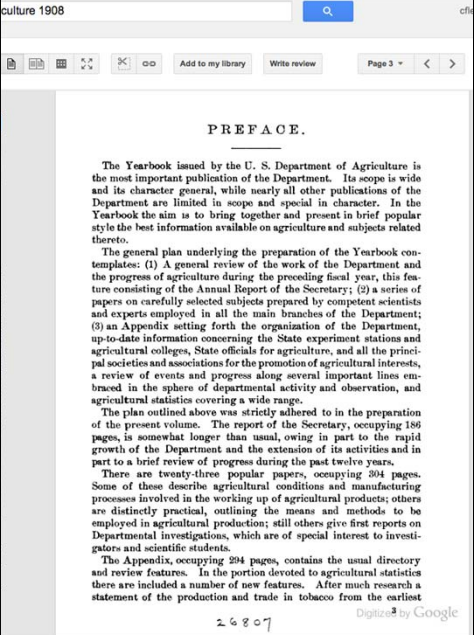
Google Example



4



Google Books search results for 'department agriculture'. The book 'Yearbook of agriculture, 1908' by United States, Dept. of Agriculture is shown with a cover featuring an owl perched on a branch. The book is available for free.



culture 1908

P R E F A C E .

The Yearbook issued by the U. S. Department of Agriculture is the most important publication of the Department. Its scope is wide and its character general, while nearly all other publications of the Department are limited in scope and special in character. In the Yearbook the aim is to bring together and present in brief popular style the best information available on agriculture and subjects related thereto.

The general plan underlying the preparation of the Yearbook contemplates: (1) A general review of the work of the Department and the progress of agriculture during the preceding fiscal year, this feature consisting of the Annual Report of the Secretary; (2) a series of papers on carefully selected subjects prepared by competent scientists and experts employed in all the main branches of the Department; (3) an Appendix setting forth the organization of the Department, up-to-date information concerning the State experiment stations and agricultural colleges, State officials for agriculture, and all the principal societies and associations for the promotion of agricultural interests, a review of events and progress along several important lines embraced in the sphere of departmental activity and observation, and agricultural statistics covering a wide range.

The plan outlined above was strictly adhered to in the preparation of the present volume. The report of the Secretary, occupying 186 pages, is somewhat longer than usual, owing in part to the rapid growth of the Department and the extension of its activities and in part to a brief review of progress during the past twelve years.

There are twenty-three popular papers, occupying 304 pages. Some of these describe agricultural conditions and manufacturing processes involved in the working up of agricultural products; others are distinctly practical, outlining the means and methods to be employed in agricultural production; still others give first reports on Departmental investigations, which are of special interest to investigators and scientific students.

The Appendix, occupying 294 pages, contains the usual directory and review features. In the portion devoted to agricultural statistics there are included a number of new features. After much research a statement of the production and trade in tobacco from the earliest

2 6 8 0 7 Digitized by Google

Books

EBOOK - FREE

Yearbook of agriculture, 1908
By United States, Dept. of Agriculture

beets Go

Showing 17 results in this book for beets - Order by: relevance | pages

Page 443 »

Pathologist in Charge of Sugar-Beet Investigations, Bureau of Plant Industry.

INTRODUCTION.

The primary object in growing sugar beets is the production of refined sugar. Any other materials, therefore, that remain or are produced in the manufacture of refined sugar from beets should be classed as by-products. These consist chiefly of beet tops (leaves and

Page 700 »

1901	48	150,759	10.34	2,007,207	484,226,430	11.09	18.3	50
1902	49	142,526	8.36	2,075,464	481,209,087	11.39	13.1	73
1903	41	216,909	8.79	1,965,812	436,811,685	11.02	14.6	58
1904	39	173,063	9.52	1,688,689	369,211,733	10.36	14.8	60

By purity coefficient is meant the percentage of sugar in the total solids of the substance tested, whether it be beets, juice, or sugar. In this table it represents the average percentage of sugar in the total solids of the beets as determined by tests made at the factories.

† Tons of 2,000 pounds each.

‡ Grouped together to avoid giving publicity to data relating to individual factories.

§ The average yield of beets per acre is found by dividing the total beets worked by the total acres harvested; the average extraction of sugar by dividing the total sugar produced by the total beets worked; the average content of sugar, coefficient of purity, and length of campaign by adding the figures reported in the average column.

Page 571 »

The imported pea moth (*Evonymia [Senasia] nigricana* Steph.) was injurious to peas at Charlevoix, Mich., this being the first record of injury in the United States.

The sugar-beet webworm (*Lasiolepe sticticalis* L.) was injurious in the vicinity of Longmont and Denver, Colo., and Hamilton City, Cal.

The southern beet webworm (*Polysiphonia lignivorella* Fab.) was observed injuring beets and spinach in southern Texas.

The beet army worm (*Ceratima crispus* Hbn.) came under observation as a pest on sugar beet, alfalfa, and peas in California; it also did injury to cucurbits. In Utah it was injurious to sugar beet.

The hop flea-beetle (*Psylliodes punctulata* Melch.), a common beet pest, was more injurious than before reported to British Columbia; it was extended to some of the

Page 445 »

It is evident, therefore, that beet tops have not received the attention due them, either as a fertilizer or as a stock food.

FULP.

The material that remains after the beets have been sliced and the sugar has been extracted is known as pulp. Fresh pulp constitutes about 80 per cent of the weight of the beets. In the process of extraction the beets lose nearly all their sugar, usually only a fraction

The screenshot shows the Google Books interface for the book "Yearbook of agriculture, 1909" by the United States Dept. of Agriculture. The book cover features an illustration of an owl perched on a branch. An overlay window titled "Books" is positioned over the right side of the page, displaying the following information:

- Books** (with "View sample" and "Add to my library" buttons)
- Buy this book** (with an "EBOOK - FREE" tag)
- Good for:**
 - ✓ Web
 - ✓ Tablet / iPad
 - ✗ eReader
 - ✗ Smartphone
- Features:**
 - ✗ Flowing text
 - ✓ Scanned pages
- Help with devices & formats
- Read the book for FREE
- Download PDF - read eReader instructions
- Learn more about books on Google Play

On the right side of the overlay, a "Price" column is partially visible, showing "No price" for several entries.

7

Library of Congress Example

The screenshot displays a digital record from the Library of Congress. At the top, there are navigation buttons for "AMERICAN MEMORY", "PREVIOUS", "NEXT", and "NEW SEARCH". The main title is "The Capital and the Bay: Narratives of Washington and the Chesapeake Bay Region, ca. 1600-1925". Below this, there is a link: "Click on picture for larger image, full item, or more versions. [Rights and Reproductions]".

The record includes a thumbnail image of the book cover titled "LIFE AND SCENES IN THE NATIONAL CAPITAL". Below the image is a "View text" link.

The main text of the record reads: "Ten years in Washington. Life and scenes in the National Capital, as a woman sees them. By Mary Clemmer Ames. Clemmer, Mary, 1839-1884. CREATED/PUBLISHED Hartford, Conn., A. D. Worthington & co. [etc., etc.] 1873. SUBJECTS Washington (D.C.)--History. Washington (D.C.)--Description and travel. MEDIUM xx, [21]-587 p. plates, ports. (incl. front.) 22 cm. CALL NUMBER F198 .A51 DIGITAL ID lhbcb 28043 http://hdl.loc.gov/loc/gdc/lhbcb 28043 http://name.umdl.umich.edu/ABE2110"

8

AMERICAN MEMORY LIBRARY OF CONGRESS

The Capital and the Bay: Narratives of Washington and the Chesapeake Bay Region, ca. 1600-1925
Ten years in Washington. Life and scenes in the National Capital, as a woman sees them.

PREVIOUS PAGE | NEXT PAGE

Back to Bibliographic Information | The Capital and the Bay Home Page | Access Historical CTFE | 2236

AMERICAN MEMORIES PREVIOUS NEXT NEW SEARCH

The Capital and the Bay: Narratives of Washington and the Chesapeake Bay Region, ca. 1600-1925

Ten years in Washington. Life and scenes in the National Capital, as a woman sees them

CHAPTER XII. WOMEN WITH CLAIMS.

Go to: [Next Section](#) | [Previous Section](#) | [Table of Contents](#)

[\[img: image\]](#)

The Senate Reception Room--The People Who Hunt It--Republican "Ladies in Waiting"--"Women with Claims"--Their Heretic Penitentiary--A Widow and Claims Agents--The Committee of Claims--A Kind-hearted Senator's Troubles--Barnabelling a Senator--A Lady of Energy--Resolved to Win--An "Office Book" a Woman--A Lady who is Feared if not Respected--Her Unfortunate Victims--Carrying "Her Measure"--The Beautiful Petitioner--The Cloudy Side of her Chivalry--Her Successes--How Government Prizes are Won.

THE room itself remains only grace, beauty and silence. The moment had not come for discussion, but I went forth without a word regarding its human aspect. To-day, dear friends, we will go in and face that. We sit down in the shadow of this Corinthian pillar, and, looking out over the most noticeable fact in that this be thought with women. A number are conversing with senators; others are gazing toward the doors which lead into the Senate. Some seem to be waiting with set faces; others are leaning back upon the sofas in attitudes of luxurious indolence. Do you ask why they are here? Are they studying the stately proportions and accessories? Not at all. It is not devotion to the aesthetic arts nor the inspiration of patriotism, which brings these women thither. They are a few, only a very few.

Page 121

[\[img: image\]](#)

women--with "claims," who, through the assistance of Congress hunt the department, the White House and the Capital.

The dejected looking woman on the sofa opposite is a widow, with numerous small children. You may be certain by the unhelpful expression of her face that she which, almost gasped and alone, she is trying to "swish through" Congress. Her home is far distant. She borrowed money to come here, she borrowed money to pay for her own board, borrowed money to pay the exorbitant fees of the claimagent, who, constantly fanning the flames of "great expectations," assures her Congress will pay her the thousands which she demands for her losses--will pay her this very session. Mentions the session is almost ended, and the widow is, with such a heavy load of debt and fear, lies hidden and forgotten in the pigeon-hole of the Committee of Claims. While it lies there, gathering dust, she is cheaply sold no longer young, and never pretty, has grown to be most burdensome to Senator--especially to the chairman of that committee. Likewise, not to be desired, in the presence of this fortune teller. No less likable to these functionaries is the sight of her hundred sisters in distress--poor or less, poor widows, with small child claims upon the Government. The chairman deems the sight of this woman and of her like. He deems it the more due he is perfectly certain that her case is not to be this session. A kind-hearted man, he is unwilling to set the seal of despair on her face by telling her the truth. She flushes a bit at last, and then remembering

Back to Bibliographic Information | The Capital and the Bay Home Page | Access Historical CTFE | 2236

PREVIOUS PAGE | NEXT PAGE

AMERICAN MEMORY LIBRARY OF CONGRESS

About the SGML version

In addition to the [full text](#) that can be viewed directly through your web browser, this document is also available encoded in Standard Generalized Markup Language (SGML) according to the [American Memory DTD](#).

No current SGML viewer is available for download, but an older, free [SGML viewer](#) is available, or the .sgm file can be downloaded and viewed with any ASCII text editor.

If you already have an SGML viewer, continue on to the [SGML version of the document](#) (1 megabytes).

Back to [previous page](#)

```
<!doctype tei2 public "-//Library of Congress - Historical Collections (American Memory)//DTD ammem.dtd//EN"
[<entity % images system "28043.ent"> %images;]>
<tei2>
<teiheader type="text" creator="National Digital Library Program, Library of Congress" status="new"
date.created="2000/10/27">
<filedesc>
<titlestmt>
<amid type="aggitemid">
lhbcb-28043
</amid>
<title>
Ten years in Washington. Life and scenes in the National Capital, as a woman sees them: a machine-readable
transcription.
</title>
<amcol>
<amcolname>
Chesapeake Bay Book Collection
</amcolname>
<amcolid type="aggid">
</amcolid>
</amcol>
<respstat>
<resp>
Selected and converted.
</resp>
<name>
American Memory, Library of Congress.
</name>
</respstat>
</titlestmt>
<publicationstmt>
<p>
Washington, DC, 2000.
</p>
</p>
```

University-plus Example: Hathi Trust

The screenshot shows the HathiTrust Digital Library homepage. At the top, there is a navigation bar with links for Home, About, Collections, Help, and Feedback. The HathiTrust logo is on the left, and a 'LOG IN' button is on the right. The main search area features a search bar with the text 'Search HathiTrust's digital library' and buttons for 'FULL-TEXT' and 'CATALOG'. Below the search bar are links for 'Advanced full-text search' and 'Search tips', and a checkbox for 'Full view only'. A question mark icon asks 'Should I search catalog or full-text?'. Below the search area, there is a paragraph describing HathiTrust as a partnership of academic and research institutions. To the right, a yellow box asks 'Want to get the most out of HathiTrust?' and provides information about partner institution accounts. At the bottom, there are four icons representing different services: 'BROWSE COLLECTIONS', 'READ BOOKS ONLINE', 'READ BOOKS ON THE GO', and 'DOWNLOAD BOOKS & CREATE COLLECTIONS'.

11

The screenshot shows a digital book viewer interface. On the left, there is a sidebar with navigation options: 'Back to catalog record', 'About this Book', 'Get this Book', 'Add to Collection', and 'Share'. The main area displays the cover of the 'National Transportation Safety Board Annual Report to Congress 1983'. The cover features a large photograph of a military aircraft on a runway. The text on the cover includes 'National Transportation Safety Board', 'Annual Report to Congress 1983', and 'TRANSPORTATION LIBRARY'. The viewer interface includes a search bar at the top right and a navigation bar at the bottom left.

12

Aviation Safety

stallation of smoke detectors in airplane lavatories.

- Require the installation of automatic fire extinguishers capable of sensing and extinguishing fires in, and adjacent to, waste receptacles in airline lavatories.
- Evaluate the electrical circuit protection—including reduced circuit breaker-rated values—needed to eliminate the potential for overheating in the lavatory flushing pump motor systems in transport category airplanes.
- Prescribe a minimum number of portable protective full face mask apparatus to be carried in the passenger compartment of airliners.

13

[← Back to Search Results](#)
← [Previous record](#) Showing record 121 of 5126 [Next record](#) →
[Cite this](#) [Export to Endnote](#)

Annual report to Congress / National Transportation Safety Board.

New Title: [Annual report](#)

Corporate Author: [United States.](#)

Language(s): English

Published: [Washington, D.C.] : National Transportation Safety Board : [1967-1984]

Subjects: [United States](#) > [National Transportation Safety Board](#) > [Periodicals](#), [Transportation > Safety > Periodicals](#), [Transportation > United States > Safety measures > Periodicals](#).

Note: Vols. for [1984] distributed to depository libraries in microfiche.

Physical Description: 17 v. : ill. ; 24-28 cm.

Locate a Print Version: [Find in a library](#)

Viewability:

- [Full view](#) 1983 (original from Northwestern University)
- [Full view](#) 1970 (original from University of Michigan)
- [Full view](#) 1968 (original from University of Michigan)
- [Full view](#) 1978-83 (original from University of Michigan)
- [Full view](#) v.2-11 1968-1977 (original from University of Michigan)
- [Full view](#) 1967 (original from Northwestern University)
- [Full view](#) 1979 (original from Northwestern University)
- [Full view](#) 1980 (original from Northwestern University)
- [Full view](#) 1982 (original from Northwestern University)

About this Book

Annual report to Congress / National Transportation ... 1983.

[View full catalog record](#)

Copyright: [Public Domain, Google-digitized](#).

Get this Book

[Find in a library](#)

[Download this page \(PDF\)](#)

[Download whole book \(PDF\)](#)

Partner login required

GPO FDSys

86 STAT. 9 - JOINT RESOLUTION DESIGNATING SEPTEMBER 17 OF EACH YEAR AS "CITIZENSHIP DAY"

Download Files	Format	PDF (289 KB)
	Descriptive Metadata	MDS
	Authenticity Metadata	FRSMB
	All Format & Metadata Files	ZIP file

Metadata

Category	Bills and Statutes
Collection	United States Statutes at Large
SubDoc Class Number	AC & 111
Publisher	U.S. Government Printing Office
Congress	86th Congress, 2nd Session, 1952
Dates in Session	Began on Tuesday, January 8, 1952 adjourned sine die on Monday, July 7, 1952
Volumes	Volume 66
Citation	66 Stat. 9
Pages	9 - 10
Law Number	Public Law 82-261
Date Approved	February 28, 1952
Full Title	Joint resolution designating September 17 of each year as "Citizenship Day"
Associated Bill Number	H.R. Res. 314

Document in Context

View Entire Statutes at Large Volume [PDF | More](#)

Vol. 66 - 82nd Congress, 2nd Session, 1952:

Part 1

Front Matter (Cover & Lists) [PDF | More](#)

Table of Contents, List of Public Laws, List of Reorganization Plans, List of Private Laws, List of Concurrent Resolutions, and List of Proclamations

Public Laws

Public Law 82-256 - An Act to provide for the withholding of certain patents that might be detrimental to the national security, and for other purposes [PDF | More](#)

Public Law 82-257 - An Act to suspend certain import duties on lead [PDF | More](#)

Public Law 82-258 - An Act to provide for the temporary free importation of zinc [PDF | More](#)

Public Law 82-259 - An Act authorizing the acquisition by the Secretary of the Interior of the Gila Pueblo, in Gila County, Ariz., for archeological laboratory and storage purposes, and for other purposes [PDF | More](#)

Public Law 82-260 - An Act to require the taking and destruction of dangerous weapons in certain cases, and for other purposes [PDF | More](#)

Public Law 82-261 - Joint resolution designating September 17 of each year as "Citizenship Day" [PDF | More](#)

Public Law 82-262 - An Act to extend to screen vehicle contractors benefits accorded star route contractors with respect to the renewal of contracts and adjustment of contract pay [PDF | More](#)

Public Law 82-263 - An Act to amend the act relating to the incorporation of Trinity College of Washington, D. C., in order to make the Archbishop of the Roman Catholic Archdiocese of Washington an ex officio member and chairman of the board of trustees of such college [PDF | More](#)

15

GPO FDSys

Document in Context

View Entire Statutes at Large Volume [PDF | More](#)

Vol. 66 - 82nd Congress, 2nd Session, 1952:

Part 1

Front Matter (Cover & Lists) [PDF | More](#)

Table of Contents, List of Public Laws, List of Reorganization Plans, List of Private Laws, List of Concurrent Resolutions, and List of Proclamations

Public Laws

Public Law 82-256 - An Act to provide for the withholding of certain patents that might be detrimental to the national security, and for other purposes [PDF | More](#)

Public Law 82-257 - An Act to suspend certain import duties on lead [PDF | More](#)

Public Law 82-258 - An Act to provide for the temporary free importation of zinc [PDF | More](#)

Public Law 82-259 - An Act authorizing the acquisition by the Secretary of the Interior of the Gila Pueblo, in Gila County, Ariz., for archeological laboratory and storage purposes, and for other purposes [PDF | More](#)

Public Law 82-260 - An Act to require the taking and destruction of dangerous weapons in certain cases, and for other purposes [PDF | More](#)

Public Law 82-261 - Joint resolution designating September 17 of each year as "Citizenship Day" [PDF | More](#)

Public Law 82-262 - An Act to extend to screen vehicle contractors benefits accorded star route contractors with respect to the renewal of contracts and adjustment of contract pay [PDF | More](#)

Public Law 82-263 - An Act to amend the act relating to the incorporation of Trinity College of Washington, D. C., in order to make the Archbishop of the Roman Catholic Archdiocese of Washington an ex officio member and chairman of the board of trustees of such college [PDF | More](#)

16

GPO FDSys

The screenshot displays the GPO FDSys interface for a document titled "66 STAT. 9 - JOINT RESOLUTION DESIGNATING SEPTEMBER 17 OF EACH YEAR AS 'CITIZENSHIP DAY'". The interface is divided into several sections:

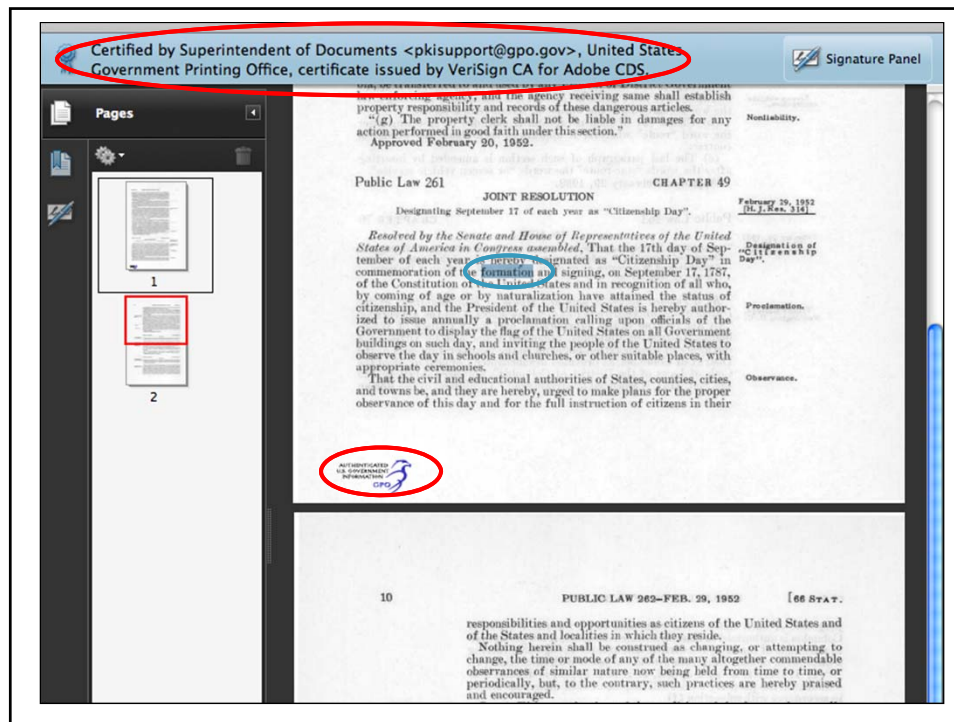
- Document in Context:** Includes "View Entire Statutes at Large Volume" and "Vol. 66 - 82nd Congress, 2nd S...".
- Download Files:** Lists "PDF (399 KB)", "MODS", "PREMIS", and "ZIP file".
- Formats:** A section for selecting document formats.
- Descriptive Metadata:** Includes "Bills and Statutes" and "United States Statutes at Large".
- Authenticity Metadata:** Includes "AE 2.111:" and "U.S. Government Printing Office".
- All Format & Metadata Files:** Includes "82nd Congress, 2nd Session, 1952" and "Begun on Tuesday, January 8, 1952 adjourned sine die on Monday, July 7, 1952".
- Metadata:** Includes "Category", "Collection", "SuDoc Class Number", "Publisher", "Congress", "Dates in Session", and "Volume".

At the bottom, there are links for "PDF I More" and a page number "17".

The screenshot shows a digital document viewer interface. At the top, it is certified by the Superintendent of Documents, United States Government Printing Office, with a signature panel for "pkisupport@gpo.gov". The document content includes:

- Public Law 261 - CHAPTER 49:** "JOINT RESOLUTION Designating September 17 of each year as 'Citizenship Day'".
- Text:** "Resolved by the Senate and House of Representatives of the United States of America in Congress assembled, That the 17th day of September of each year hereby designated as 'Citizenship Day' in commemoration of the formation and signing, on September 17, 1787, of the Constitution of the United States and in recognition of all who, by coming of age or by naturalization have attained the status of citizenship, and the President of the United States is hereby authorized to issue annually a proclamation calling upon officials of the Government to display the flag of the United States on all Government buildings on such day, and inviting the people of the United States to observe the day in schools and churches, or other suitable places, with appropriate ceremonies."
- Text:** "That the civil and educational authorities of States, counties, cities, and towns be, and they are hereby, urged to make plans for the proper observance of this day and for the full instruction of citizens in their responsibilities and opportunities as citizens of the United States and of the States and localities in which they reside."

The interface includes a "Pages" sidebar on the left with thumbnails for pages 1 and 2, and a "Signature Panel" on the right. The document is signed and dated "February 20, 1952 [St. J. Res. 316]".



Ingredients

- Underlying materials: the basic inputs
 - Catalog record or equivalent
 - Set of images
 - Marked-up text
- Secondary elements
 - Additional output versions
 - PDFs, other eBooks, handoff to hard-copy producers, PKI
- Elements from “presentation” IT systems
 - Interface, presentation structure
 - Index and search functions
 - Additional output versions may be produced on the fly


Typical Production

- Initial stage
 - **Cataloging**
 - **Scanning to produce images**
- Second stage
 - **OCR the images to produce text**
- Third stage
 - **Process the text to added markup, other prep**
 - **Create some additional output versions (PDF, eBook, hard copy)**
- Fourth stage
 - **Load elements, index data, etc.**
 - **Create some additional output versions**


Typical Production **Standards or Guidelines**

- Initial stage
 - **Cataloging**
 - **library standards: MARC, MODS, etc.**
 - **Scanning to produce images**
 - **FADGI guidelines**
- Second stage
 - **OCR the images to produce text**
 - **various software apps, effectiveness has increased over time**
- Third stage
 - **Process the text to added markup, other prep**
 - **practices vary (TEI Lite, AM DTD, METS, other)**
 - **Create some additional output versions (PDF, eBook, hard copy)**
 - **practices vary**
- Fourth stage
 - **Load elements, index data, etc.**
 - **local option**
 - **Create some additional output versions**
 - **practices vary**

Typical Production Standards or Guidelines

- 
- Second stage
 - OCR the images to produce text
 - various software apps, effectiveness has increased over time
 - Third stage
 - Process the text to added markup, other prep
 - practices vary (TEI Lite, AM DTD, METS, other)
 - Create some additional output versions (PDF, eBook, hard copy)
 - practices vary
 - Fourth stage
 - Load elements, index data, etc.
 - local option
 - Create some additional output versions
 - practices vary

Typical Production Standards or Guidelines

- Initial stage
 - Cataloging
 - library standards: MARC, MODS, etc.
 - Scanning to produce images
 - FADGI guidelines
- 

Typical Production Standards or Guidelines

- Initial stage
 - **Cataloging**
 - library standards: MARC, MODS, etc.
 - **Scanning to produce images**
 - FADGI guidelines



- Third stage
 - **Process the text to added markup, other prep**
 - practices vary (TEI Lite, AM DTD, METS, other)
 - **Create some additional output versions (PDF, eBook, hard copy)**
 - practices vary
- Fourth stage
 - **Load elements, index data, etc.**
 - local option
 - **Create some additional output versions**
 - practices vary

Typical Production Standards or Guidelines

- Initial stage
 - **Cataloging**
 - library standards: MARC, MODS, etc.
 - **Scanning to produce images**
 - FADGI guidelines
- Second stage
 - **OCR the images to produce text**
 - various software apps, effectiveness has increased over time
- Third stage
 - **Process the text to added markup, other prep**
 - practices vary (TEI Lite, AM DTD, METS, other)
 - **Create some additional output versions (PDF, eBook, hard copy)**
 - practices vary
- Fourth stage
 - **Load elements, index data, etc.**
 - local option
 - **Create some additional output versions**
 - practices vary

The screenshot shows the homepage of the Federal Agencies Digitization Guidelines Initiative. The navigation bar includes links for Home, About, Guidelines, Resources, Review Process, and News & Events. The main content area is divided into several sections: 'About This Initiative', 'Digitization Guidelines', 'News & Events', and 'Resources'. The 'About This Initiative' section contains three sub-sections: 'Still Image Working Group', 'Audio-Visual Working Group', and 'Audio Digitization System Performance'. The 'Still Image Working Group' section is circled in red. Below the screenshot, the URL <http://www.digitizationguidelines.gov/> is provided.

27

18 Participating Agencies
<http://www.digitizationguidelines.gov/participants/>

Often participating, not "official": NASA, NOAA, National Museum of Health and Medicine (U.S. Army), U.S. Supreme Court

2

Scanned Image Specifications

- Something of a line-drawing exercise
- Purpose, goal
 - If the term *preservation* is used, the stakes are higher
- Category of original item
 - If materials are rare, or only held in a few libraries, the stakes are higher
 - If materials contain important illustration or diagrams (and in color), the stakes are higher

How do you indicate higher stakes?

- FADGI guidelines use *stars*
- Every specification factor has an *aim point*
 - For example, *spatial resolution might have 400 ppi as the aim point*
- Every aim point has a way to measure
 - *Every measurement allows for plus-minus tolerances*
- Star rating
 - 4 star – very close to aim point
 - 3 star – pretty close
 - 2 star – not bad
 - 1 star – just OK


Rough Rule of Thumb

- 4-star
 - Artwork, world and national treasures, etc.
- 3-star
 - Unique manuscript and archival documents
- 2-star
 - General collections (not unique) and printed publications
- 1-star
 - Text conversion, image quality not important



Recent LC Example 1

- High end, set of important books, foreign imprints
- Non-destructive scanning (“don’t hurt the books”)
- 400 ppi and 24-bit RGB color
- 120 or 180 degree *V* opening in a book cradle
- Overhead scanner shall meet a **3-star or better imaging performance** rating as defined in the FADGI 2010 guidelines
- Scanner calibration and image processing
 - to be set for originals with a density range of 0.04 to 2.04 and the following aimpoints - a medium density range appropriate for a variety of originals (printed text and graphics, some photos) and optimized for general monitor viewing on monitors calibrated to sRGB or AdobeRGB 1998 environments:
 - Density of 0.04 (patch 10 on the DICE targets, patch A on Q13/Q14) should be RGB level 242-242-242
 - Density of 0.75 (patch 16 on the DICE targets, patch M on Q13/Q14) should be RGB level 104-104-104
 - Density of 2.04 (patch 20 on the DICE targets, patch 19 on Q13/Q14) should be RGB level 11-11-11

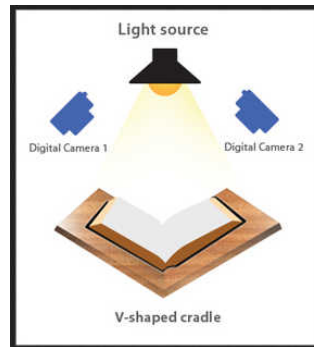
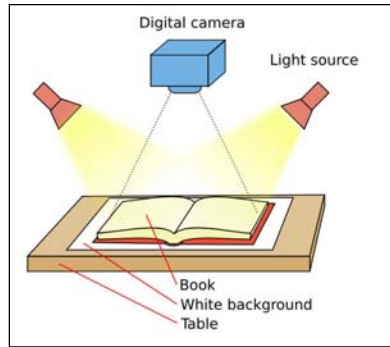
Recent LC Example 1

- High end, set of important books, foreign imprints
- Non-destructive scanning (“don’t hurt the books”)
- 
- 120 or 180 degree V opening in a book cradle
- Overhead scanner shall meet a **3-star or better imaging performance** rating as defined in the FADGI 2010 guidelines
- Scanner calibration and image processing
 - to be set for originals with a density range of 0.04 to 2.04 and the following aimpoints - a medium density range appropriate for a variety of originals (printed text and graphics, some photos) and optimized for general monitor viewing on monitors calibrated to sRGB or AdobeRGB 1998 environments:
 - Density of 0.04 (patch 10 on the DICE targets, patch A on Q13/Q14) should be RGB level 242-242-242
 - Density of 0.75 (patch 16 on the DICE targets, patch M on Q13/Q14) should be RGB level 104-104-104
 - Density of 2.04 (patch 20 on the DICE targets, patch 19 on Q13/Q14) should be RGB level 11-11-11

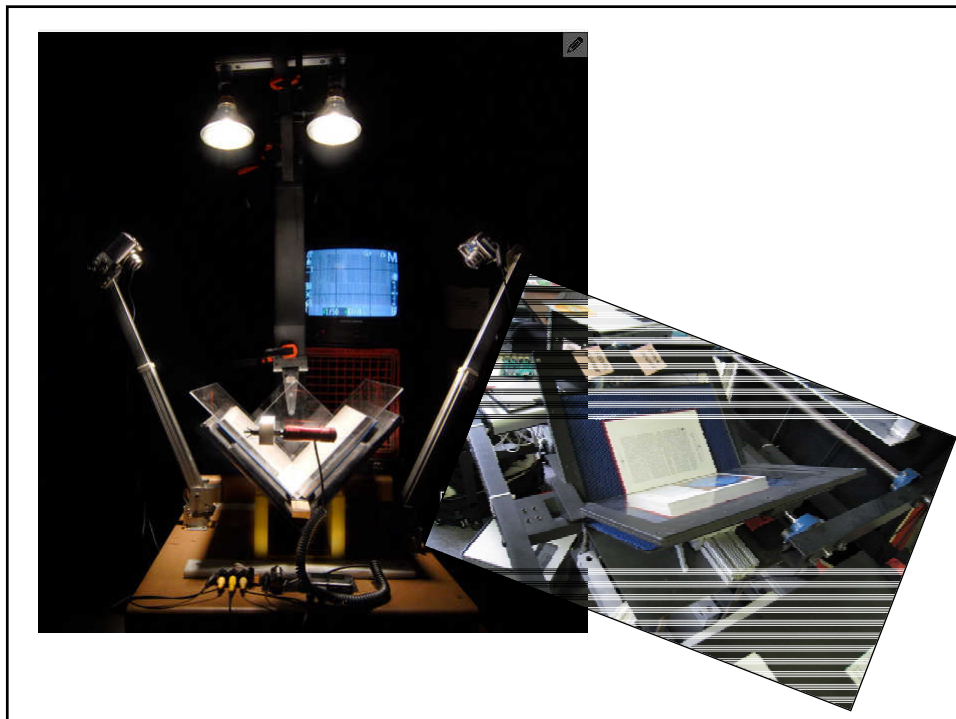
Recent LC Example 1

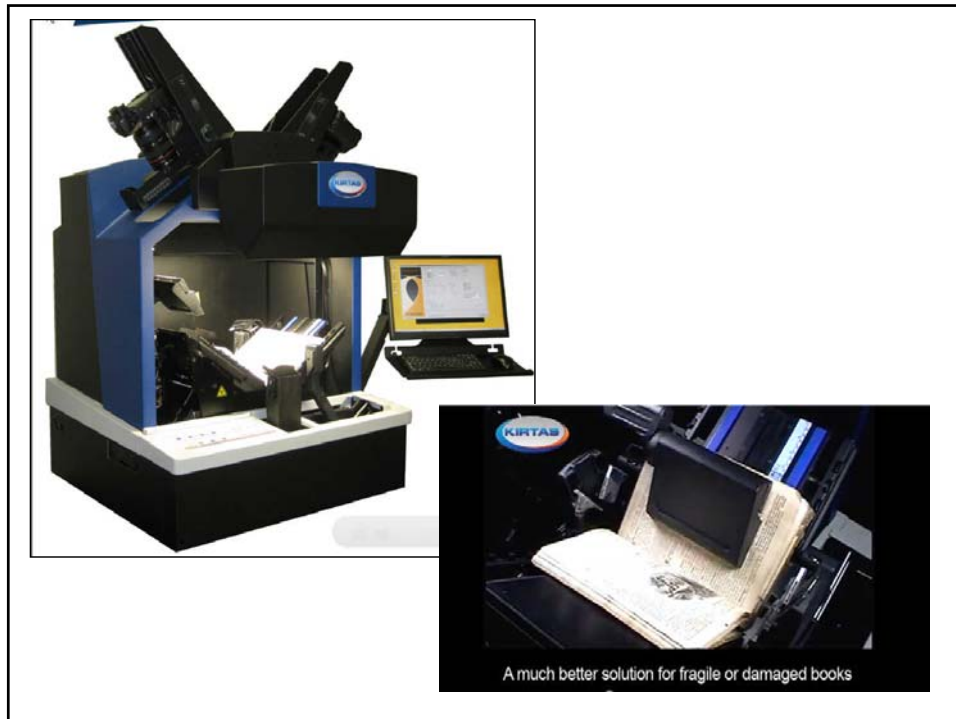
- High end, set of important books, foreign imprints
- Non-destructive scanning (“don’t hurt the books”)
- 
- 
- guidelines
- Scanner calibration and image processing
 - to be set for originals with a density range of 0.04 to 2.04 and the following aimpoints - a medium density range appropriate for a variety of originals (printed text and graphics, some photos) and optimized for general monitor viewing on monitors calibrated to sRGB or AdobeRGB 1998 environments:
 - Density of 0.04 (patch 10 on the DICE targets, patch A on Q13/Q14) should be RGB level 242-242-242
 - Density of 0.75 (patch 16 on the DICE targets, patch M on Q13/Q14) should be RGB level 104-104-104
 - Density of 2.04 (patch 20 on the DICE targets, patch 19 on Q13/Q14) should be RGB level 11-11-11

Non-Destructive Scanning



Drawings from Wikipedia





Recent LC Example 2

- Congressional documents, widely held, recent and not rare, no illustrations
- Documents will be disbound, can be sheet-fed
 - aka Automatic Document Feeder (ADF)
 - In some contexts, especially if the books are guillotined to free up the pages (ours are not), this would be called *destructive* scanning.
- SOW calls for 400 ppi and 8-bit grayscale
- Far fewer other specifications
- In a task order where 2-star is understood
 - *BTW: still uncompressed image data in TIFF wrapper, but studying JPEG 2000, even lossy for some content*

Recent LC Example 2

- Congressional documents, widely held, recent and not rare, no illustrations



– *BTW: still uncompressed image data in TIFF wrapper, but studying JPEG 2000, even lossy for some content*

Recent LC Example 2

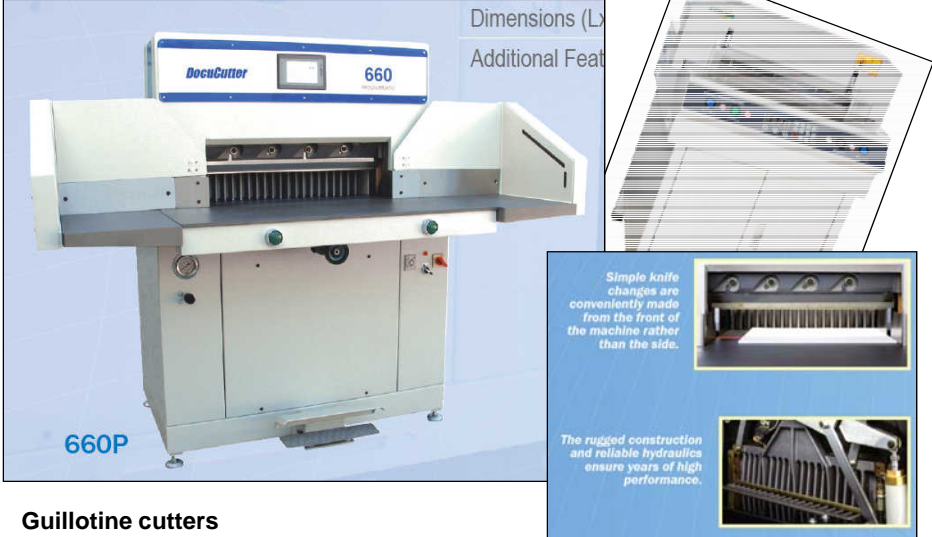
- Congressional documents, widely held, recent and not rare, no illustrations



File Format

- TIFF file wrapper
- RGB 24-bit color or grayscale
- LC produces uncompressed images
- Other organizations use lossless LZW compression
- However, file size is leading us to consider JPEG 2000
- JPEG 2000 can be used in lossless or lossy modes

Destructive Scanning



Dimensions (L)

Additional Feat

DocuCutter 660

660P

Guillotine cutters

Simple knife changes are conveniently made from the front of the machine rather than the side.

The rugged construction and reliable hydraulics ensure years of high performance.



FADGI Imaging Guideline

Still image specifications – this is what we all “used to do”

- color/monochromatic
- pixels per inch (ppi; aka old “dpi”)
- bit depth
- . . . usually output-referred

We want to move toward more scientific metrics

<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 • Delta E₂₀₀₀ • 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

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)

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

From this document:
<http://www.digitizationguidelines.gov/guidelines/DIFfinal.pdf>

Rethinking Resolution

- We have all had the tendency to talk about the pixels per inch (ppi) delivered by a scanner as the *resolution*
- Regarding the scanner setting, it would be better to call it *SAMPLING FREQUENCY* or *SAMPLING RATE*.

Rethinking Resolution

- Meanwhile, there is an ISO standard for *RESOLUTION* (ISO 12233). This gets at the actual “optical” result delivered by the device.
- Among other elements, the ISO standard asks you to measure resolution via *SPATIAL FREQUENCY RESPONSE (SFR)*. You image a target, do some math, and present the numbers on a graph.

Rethinking Resolution

- Knowing the Spatial Frequency Response (SFR) lets you compare the scanner's setting (sampling frequency) with what you get (in true optical terms).
- FADGI documents call what you get *EFFECTIVE RESOLUTION*.
- The comparison of scanner setting and effective resolution is what we call *SAMPLING EFFICIENCY*.

Sampling Efficiency and Corresponding Effective Resolution

Desired Spatial Resolution	Scanner Sampling Efficiency and Corresponding Effective Resolution											
	100%	96%	95%	91%	90%	85%	81%	80%	75%	70%	65%	60%
5000 ppi	5000	4800	4750	4550	4500	4250	4050	4000	3750	3500	3250	3000
2500 ppi	2500	2400	2375	2275	2250	2125	2025	2000	1875	1750	1625	1500
1200 ppi	1200	1152	1140	1092	1080	1020	972	960	900	840	780	720
800 ppi	800	768	760	728	720	680	648	640	600	560	520	480
600 ppi	600	576	570	546	540	510	486	480	450	420	390	360
400 ppi	400	384	380	364	360	340	324	320	300	280	260	240
300 ppi	300	288	285	273	270	255	243	240	225	210	195	180
200 ppi	200	192	190	182	180	170	162	160	150	140	130	120
100 ppi	100	96	95	91	90	85	81	85	75	70	65	60
Spurious Resolution and Storage Overhead*		108%	111%	121%	123%	138%	152%	156%	178%	204%	237%	278%
Imaging Performance Level**	4-star (> 95% and up to 100%)		3-star (> 90% and up to 95%)		2-star (> 80% and up to 90%)			1-star (80% or less)				

Sampling Efficiency and Corresponding Effective Resolution

Desired Spatial Resolution	100%	96%	95%	91%	90%	85%	81%	80%	75%	70%	65%	60%
5000 ppi	5000	4800	4750	4550	4500	4250	4050	4000	3750	3500	3250	3000
2500 ppi	2500	2400	2375	2275	2250	2125	2025	2000	1875	1750	1625	1500
1200 ppi	1200	1152	1140	1092	1080	1020	972	960	900	840	780	720
800 ppi	800	768	760	728	720	680	648	640	600	560	520	480
600 ppi	600	576	570	546	540	510	486	480	450	420	390	360
400 ppi	400	384	380	364	360	340	324	320	300	280	260	240
300 ppi	300	288	285	273	270	255	243	240	225	210	195	180
200 ppi	200	192	190	182	180	170	162	160	150	140	130	120
100 ppi	100	96	95	91	90	85	81	85	75	70	65	60
Spurious Resolution and Storage Overhead*		108%	111%	121%	123%	138%	152%	156%	178%	204%	237%	278%
Imaging Performance Level**	4-star (> 95% and up to 100%)		3-star (> 90% and up to 95%)		2-star (> 80% and up to 90%)			1-star (80% or less)				

Sampling Efficiency and Corresponding Effective Resolution

Desired Spatial Resolution	100%	96%	95%	91%	90%	85%	81%	80%	75%	70%	65%	60%
5000 ppi	5000	4800	4750	4550	4500	4250	4050	4000	3750	3500	3250	3000
2500 ppi	2500	2400	2375	2275	2250	2125	2025	2000	1875	1750	1625	1500
1200 ppi	1200	1152	1140	1092	1080	1020	972	960	900	840	780	720
800 ppi	800	768	760	728	720	680	648	640	600	560	520	480
600 ppi	600	576	570	546	540	510	486	480	450	420	390	360
400 ppi	400	384	380	364	360	340	324	320	300	280	260	240
300 ppi	300	288	285	273	270	255	243	240	225	210	195	180
200 ppi	200	192	190	182	180	170	162	160	150	140	130	120
100 ppi	100	96	95	91	90	85	81	85	75	70	65	60
Spurious Resolution and Storage Overhead*		108%	111%	121%	123%	138%	152%	156%	178%	204%	237%	278%
Imaging Performance Level**	4-star (> 95% and up to 100%)		3-star (> 90% and up to 95%)		2-star (> 80% and up to 90%)			1-star (80% or less)				

4-star: Likely to require a slower and more costly process.

Sampling Efficiency and Corresponding Effective Resolution

Desired Spatial Resolution	Scanner Sampling Efficiency and Corresponding Effective Resolution											
	100%	96%	95%	91%	90%	85%	81%	80%	75%	70%	65%	60%
5000 ppi	5000	4800	4750	4550	4500	4250	4050	4000	3750	3500	3250	3000
2500 ppi	2500	2400	2375	2275	2250	2125	2025	2000	1875	1750	1625	1500
1200 ppi	1200	1152	1140	1092	1080	1020	972	960	900	840	780	720
800 ppi	800	768	760	728	720	680	648	640	600	560	520	480
600 ppi	600	576	570	546	540	510	486	480	450	420	390	360
400 ppi	400	384	380	364	360	340	324	320	300	280	260	240
300 ppi	300	288	285	273	270	255	243	240	225	210	195	180
200 ppi	200	192	190	182	180	170	162	160	150	140	130	120
100 ppi	100	96	95	91	90	85	81	85	75	70	65	60
Spurious Resolution and Storage Overhead*		108%	111%	121%	123%	138%	152%	156%	178%	204%	237%	278%
Imaging Performance Level**	4-star (> 95% and up to 100%)		3-star (> 90% and up to 95%)		2-star (> 80% and up to 90%)			1-star (80% or less)				

2-star: Likely to permit a faster and less costly process.

Rethinking Resolution

Thanks to Barry Wheeler for his very helpful Signal blogs:

<http://blogs.loc.gov/digitalpreservation/2012/12/what-resolution-should-i-use-part-1/>

<http://blogs.loc.gov/digitalpreservation/2013/01/what-resolution-should-i-use-part-2/>

<http://blogs.loc.gov/digitalpreservation/2013/03/what-resolution-should-i-use-part-3/>

From the revised guideline

Performance Rating		
*	**	***
← Less Critical - More Critical →		

Sampling Frequency		
PERFORMANCE LEVEL	AIM	TOLERANCE
****	User selected (e.g., 300, 400, 600 dpi)	+/- 0.50% of AIM
***		+/- 0.75%
**		
*		

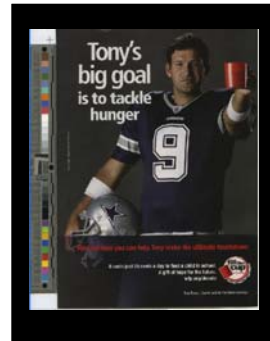
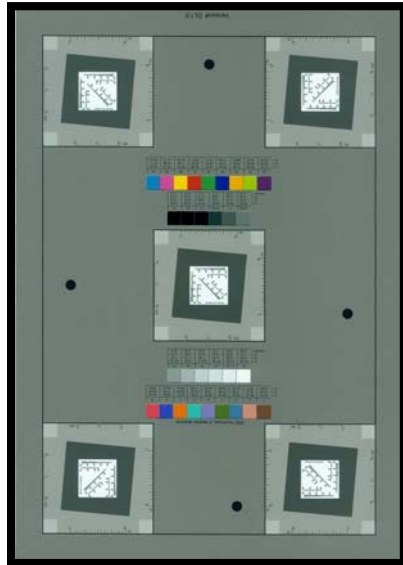
Resolution - Spatial Frequency Response (SFR) High Frequency (note 1) (native response, no sharpening)			
PERFORMANCE LEVEL	AIM	TOLERANCE (for both horizontal and vertical directions equally)	
		Lower Limit sampling efficiency (luminance)	Upper Limit SFR response at half sampling frequency
****	100% Efficiency and 10% SFR	> 95%	< 0.20
***		> 90%	< 0.30
**		> 80%	< 0.40
*		< 80%	> 0.40

http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf

Tools to Support Image Performance Measurement

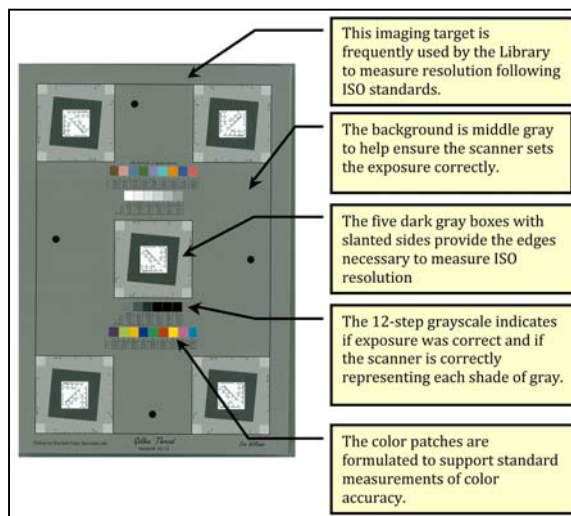
- Digital Image Conformance Evaluation (DICE) System
 - **Device Target** – Imaging Device Performance
 - **Object Target** – Actual Image Quality
 - **Software** for Evaluation/Validation
- Software still under development
 - Built on a commercial library of math tools
 - Future: FADGI plan to build as open source
- Commercial equivalent sold by Image Science Associates, Inc.

Device and Object Targets

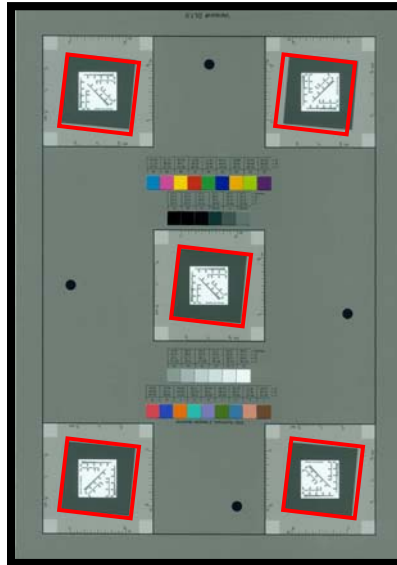


Object target as positioned for use

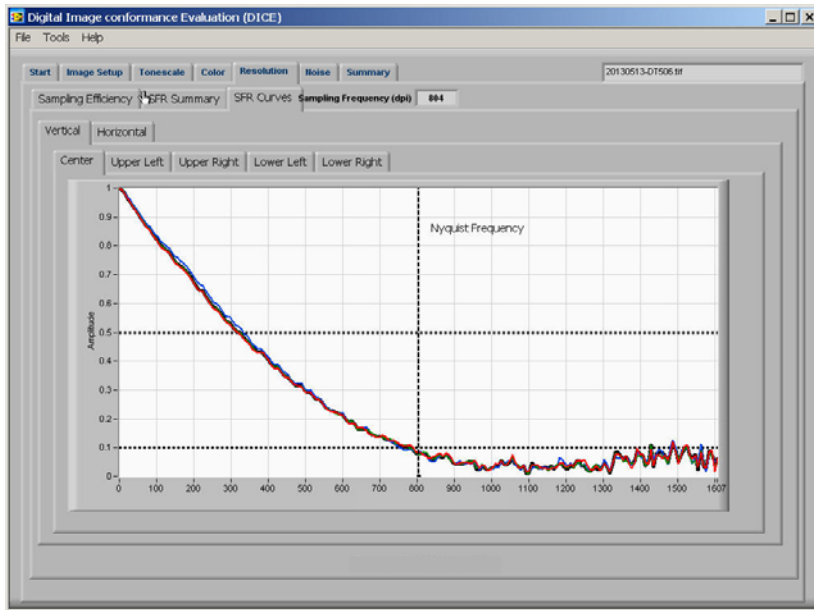
Device Target



Device Target



DICE – SFR detail page



**Federal Agencies
Digitization Guidelines Initiative**

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

Carl Fleischhauer cfle@loc.gov