



Digital Imaging 101

Converting Tangible Publications to Digital Assets

David Walls, Preservation Librarian
U.S. Government Publishing Office, March 20, 2017

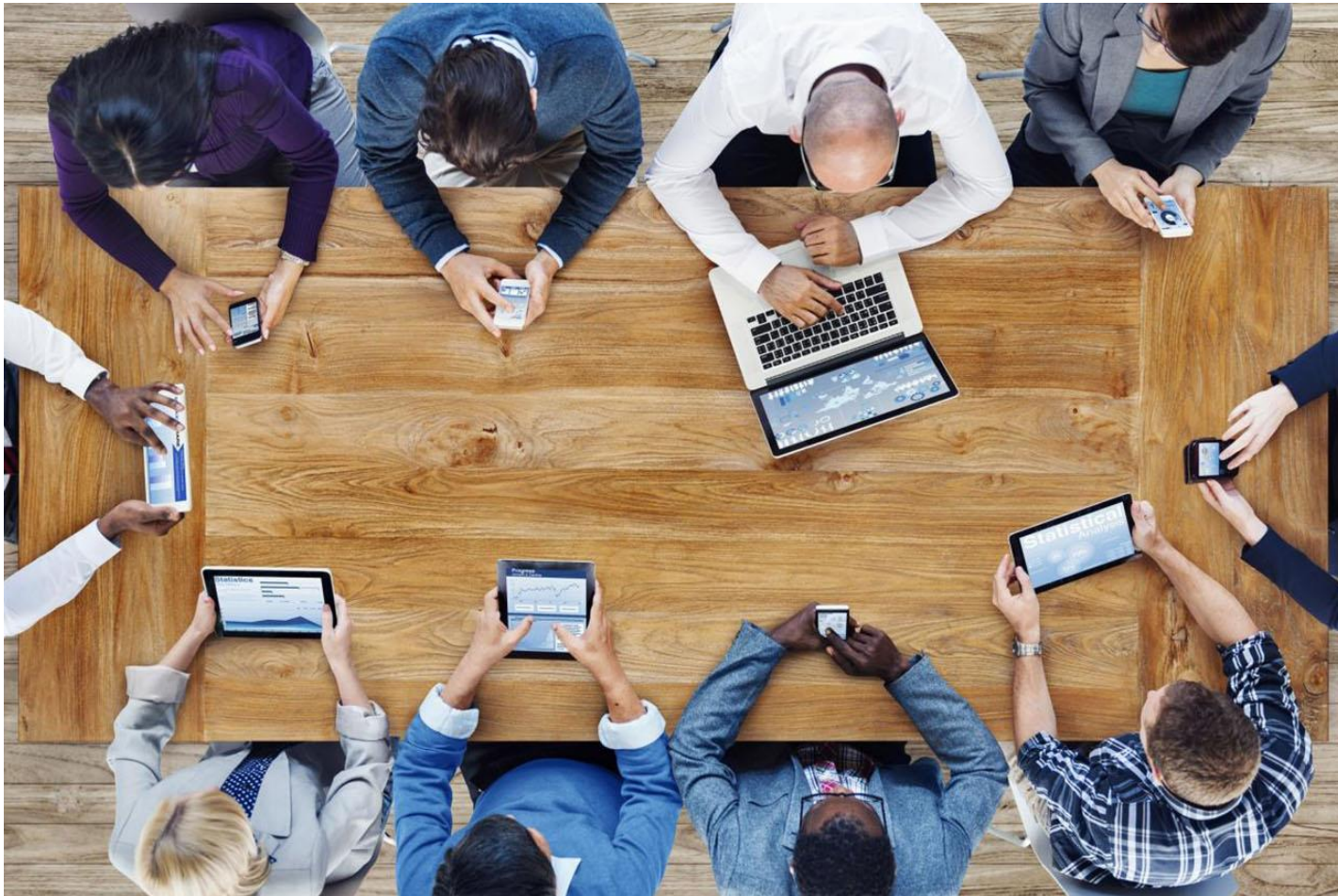
Librarians

Acquire, manage, preserve and provide access to a set of information assets relied upon by our information customers.

Connecting Content

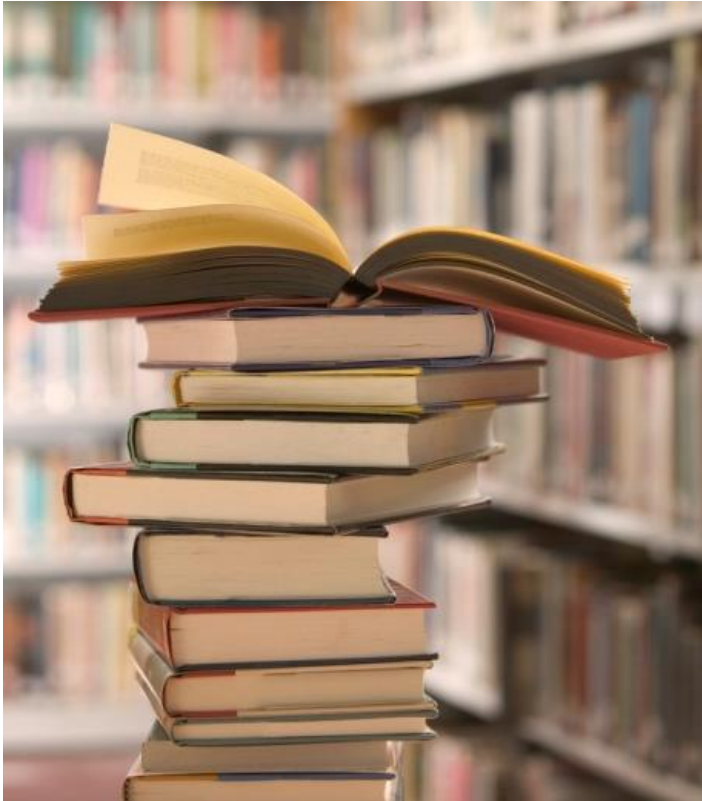


Community of Users





Why digitize? Access



- Paper collections are location specific
- One user at a time
- Limited online catalog records before 1976
- No cross content search capability
- Collections are overwhelming available space.
- Divide between paper and digital content widening
- Brings collection materials up to date with modern discovery and search methods
- Some formats becoming obsolete

Why digitize? Preservation

- Older content is fragile due to acidic decay
- Content is vulnerable to fire, flood, pests, environmental disasters, people who are too lazy to photocopy, and other acts of vandalism
- Primary Source and other special collection materials are unique and vulnerable
- Special Collections less accessible due to preservation and security concerns
- No way to backup content to maintain continuity of access
- Collections stored out of sight become collections stored out of mind
- Some formats becoming obsolete





Selecting content for digitization: access

- Library mission statement
- Collection Development Plans, Priorities
- Usage statistics from collection items
- User driven selection through reference contacts and social networking
- Professional experience
- Professional networking



Planning for digitization

- What are you going to digitize and why?
- Can you digitize it? (copyright)
- How will you get support for your digital project?
- How are you going to digitize the, items, collection?
- What preservation files do you need?
- What access derivatives do you need for the user experience?
- Where will you store your preservation files?
- Who will host your digital content?
- How will your users find your digitized content?
- How will you share progress with stakeholders?



Copyright

- www.copyright.gov
- <http://librarycopyright.net/>



Insourcing vs. outsourcing

- Adding digitization duties to existing staff
- Setting up a digital workspace
- The essentials: consistent training for staff and maintenance and upgrading for equipment
- How much can you do?
- Tracking costs, including overhead costs.
- Vendors keep consistent with technology. It's business!
- Vendors will only do what the contract tells them to.
- A pilot project is essential to test your specs and your vendors ability to work with you and deliver what you want.



Digitization Best Practices

- The digitization process will preserve the look of the original source item
- The source publication must contain all of the information content intended by its publisher.
- All information content from the original source item must be captured in the digitization process.
- The physical condition of the source publication must not compromise the image capture or OCR software
- Gutter margins must be wide enough to digitize the source publication without excessive page curvature.
- The chosen file formats must be preservable and conform to user access requirements



Preservation and access files

Preservation formats:

- Tiff
- JPEG2000
- PNG
- PDF/A

Access formats:

- PDF/A
- JPEG
- JPEG2000
- e-Pub
- Kindle
- DAISY

Contributing Partner Guidance



Guidance for Contributing Digital Content to FDsys/govinfo

The following guidance for digital reformatting is based on current best practices. The results will provide the best image capture for the range of information products to be ingested into FDsys/govinfo.

The guiding principle for digitization rests in GPO's definition of preservation copy of record for digital content:

The preservation copy of record for digital content is the preservation master file stored in a trustworthy repository. Derivatives of the preservation master copy are made available for access. *The digital copy of record should be produced to specifications that will allow the creation of a printed facsimile version, should one be needed.*

PRESCRIBED TECHNICAL SPECIFICATIONS

Specifications for preservation master, access derivative files, and metadata are described as "Preferred" and "Acceptable." The Preferred specifications follow the latest best practices and guidance for digital preservation and support GPO's operation of FDsys/govinfo as a digital repository. The Acceptable specifications are given to accommodate a variety of stakeholders and their technical capability.

Digitize publications using the preferred JPEG 2000, or the acceptable TIFF master file format following the prescribed sampling rate of 300 to 600 ppi to accurately capture the original content. Each page scanned will result in a separate digital file in the prescribed master format. An access file in pdf format will be derived from the master image files. Optical character recognition software will be run to embed machine readable text into the pdf.

	PREFERRED	ACCEPTABLE
MASTER FILE	Uncompressed JPEG 2000, 300-600 ppi, 24 bit RGB color conforming to the ISO/IEC 15444-1 standard for JPEG 2000	Uncompressed TIFF 6.0, 300-600 ppi, 24 bit RGB color
ACCESS FILE	PDF/A 2-b with embedded Optical Character Recognition	PDF/A with embedded Optical Character Recognition
TECHNICAL METADATA	MIX XML*	
BIBLIOGRAPHIC METADATA	MARC XML	MARC XML

* NISO z39.87 defines a set of metadata elements for raster images to enable users to develop, exchange, and interpret digital image files. These elements, such as information about the compression, color profile, resolution, scanner or digital camera make and model, can be recorded and preserved as technical information for still images. When possible, this information should be recorded as a set of XML elements conforming to the NISO metadata for images in xml Schema (MIX).

- Increases public access to legacy collections
- Provides digital surrogates of “Copies of Record.”
- Preserves digital copies of Federal government publications in a Trustworthy Digital Repository: FDsys/govinfo

Digitization specifications

www.digitizationguidelines.gov

[Glossary](#) | [Participants](#) | [Contact Us](#)

Search this site

 **Federal Agencies
Digital Guidelines Initiative**

[Submit comments](#) on guidelines or on our digitization efforts in general or
[Tweet #FADGI](#)

[Home](#) | [About](#) | [Guidelines](#) | [Resources](#) | [Review Process](#) | [News & Events](#)

[Audio System Performance](#) | [Content Categories & Digitization Objectives](#) | [Digital Imaging Framework](#) | [Digitizing Motion Picture Film](#) | [Digitization Activities - Project Planning](#) | [Embedded Metadata in Broadcast WAVE Files](#) | [Embedded Metadata in Digital Still Images](#) | [Embedded Metadata in DPX Files](#) | [Embedded Metadata in TIFF Images](#) | [File Format Comparisons](#) | [MXF Application Specification](#) | [Technical Guidelines for Digitizing Cultural Heritage Materials](#)

[Home](#) > [Guidelines](#) > [Technical Guidelines for Digitizing Cultural Heritage Materials](#)

 Print  Subscribe  Share/Save

Guidelines: Technical Guidelines for Digitizing Cultural Heritage Materials

Still Image Working Group

The document below, revised by the Still Image Working Group in 2015 and approved in September 2016, is an update of the 2010 *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*. This revision of the guidelines expands on the earlier works and incorporates new material reflecting the advances in imaging science and cultural heritage imaging best practice. The format has changed. Relevant information specific to defined imaging tasks has been grouped into individual chapters, along with guidance on how to apply the guidelines to specific tasks. These guidelines are intended to be used in conjunction with DICE, Golden Thread, [OpenDICE and AutoSFR](#) (Digital Image Conformance Environment) targets and software. Together, these guidelines and the DICE testing and monitoring system provide the foundation for a FADGI-compliant digitization program.

[Download OpenDice and AutoSFR](#) open source software and manuals.

Working Groups

[Still Image Working Group](#)

This group is involved in a cooperative effort to develop common digitization guidelines for still image materials.

[Audio-Visual Working Group](#)

This group works collaboratively on common and sustainable technical guidelines, methods, and practices for

<http://www.loc.gov/standards/marcxml/>

[The Library of Congress](#) >> [Standards](#) >> [MARC](#)

MARCXML MARC 21 XML Schema Official Web Site

The Library of Congress' Network Development and MARC Standards Office is developing a framework for working with MARC data in a XML environment. This framework is intended to be flexible and extensible to allow users to work with MARC data in ways specific to their needs. The framework itself includes many components such as schemas, stylesheets, and software tools.

MARCXML Architecture

- [MARCXML Design Considerations](#)
- [MARCXML Architecture](#)
- [MARCXML Uses and Features](#)

MARCXML Schema & Documentation

- [MARCXML Schema](#)
- [MARCXML Illustrated](#)

MARCXML Example Documents

- Example MARCXML Documents
 - Carl Sandburg's Arithmetic (single record)
 - [Original MARC \(2709\) Record](#)
 - [MARCXML Instance](#)
 - [HTML Presentation](#)
 - [MODS Transformation](#)
 - [Dublin Core Transformation](#)
 - Example of a collection (multiple records)
 - [MARCXML Instance](#)

Content and access systems

Local repository

OCLC Digital Archive/ContentDM

Rosetta

LOCKSS

Drupal

FDsys

Internet Archive/Wayback

Hathi Trust



Information sources:

List servs:

digipres@ala.org

Training:

www.nedcc.org

www.lyrasis.org

Guidelines

www.digitizationguidelines.gov

<http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>



A successful digitization project:

- Results from careful planning
- Supports the library's mission, collection development, and preservation priorities
- Provides an enhanced experience for content users
- Based on realistic goals
- Increases institutional knowledge
- Preserves tangible content
- Connects content to community



David Walls
Preservation Librarian, GPO
dwalls@gpo.gov