Digital Imaging 101

Converting Tangible Publications
to Digital Assets

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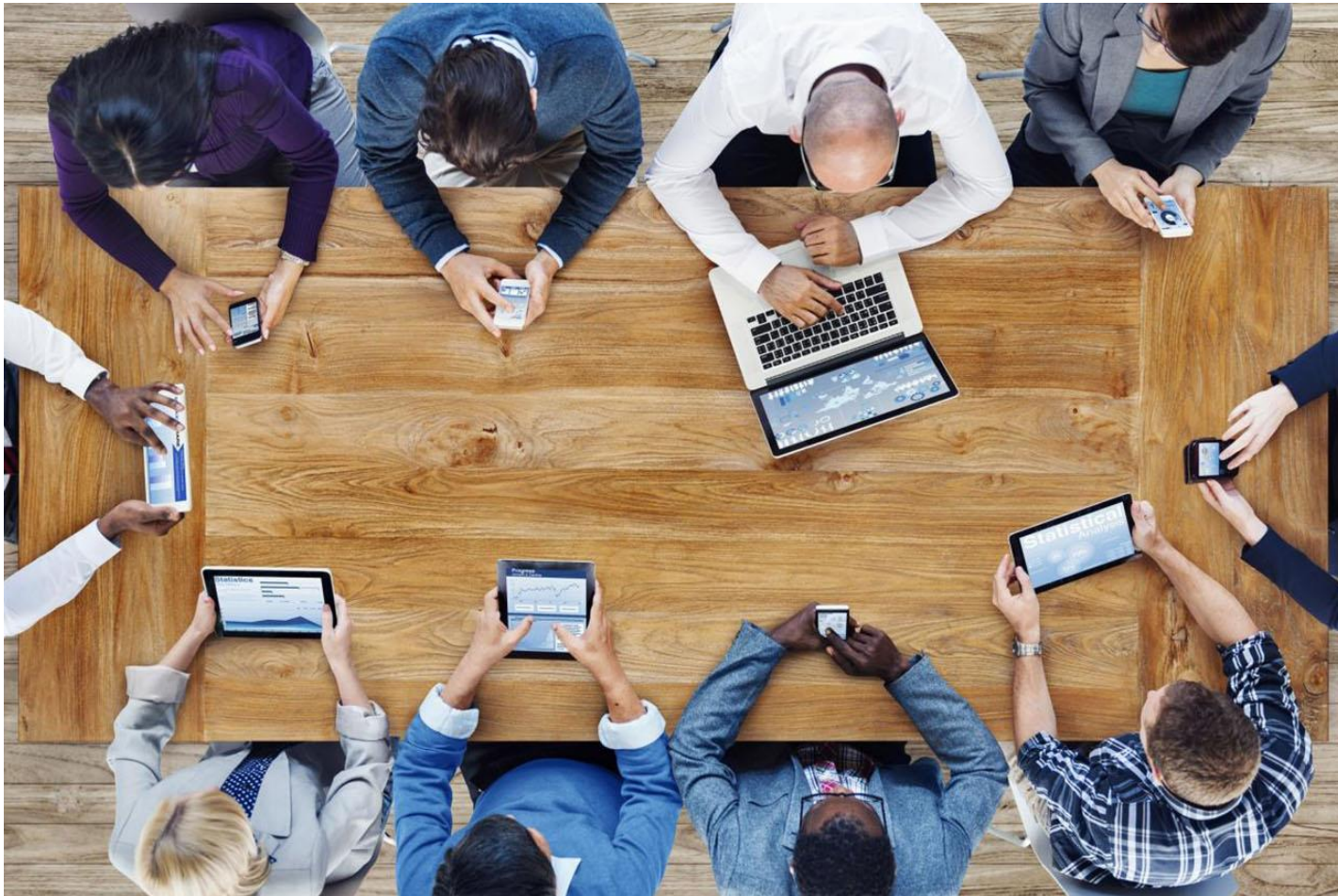U.S. Government Publishing Office, March 20, 2017**

Librarians

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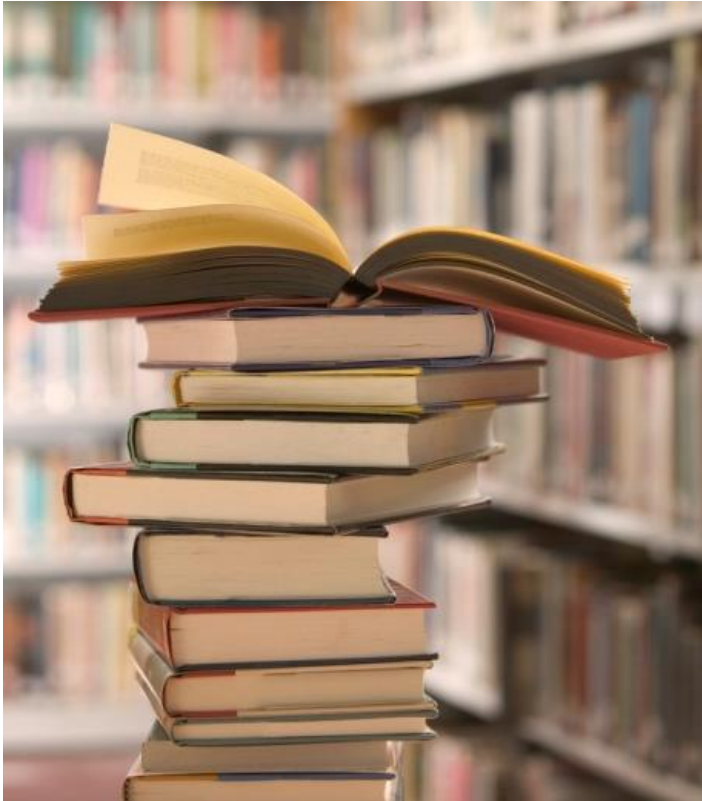
Connecting Content

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 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

- 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

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