Digitization on a Shoestring

Part 2

GALILEO User Conference 2020: DLG Track
June 12, 2020

Mary Willoughby
Digital Conversion and Curation Librarian, Digital Library of Georgia

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC LIBRARY SERVICE

Georgia's Virtual Library
GALILEO
An Initiative of the University System of Georgia

Planning for digital conversion

Digitization on a Shoestring Part 1 https://bit.ly/shoestring1

What are the goals of your project?

There is no one right answer for digital conversion.

Think about your short term, mid term, and long term goals-- what does success look like?

Which of these timescales is most important to your stakeholders, internal and external?

One size does not fit all, and more isn't always better.

What are the goals of your project?

Align digitization processes with project goals.

- You can be successful in providing access with lower resolution scans, but be realistic about your expectations for reuse.
  - Scans made with less sophisticated equipment preserve less detail and provide less accurate color
  - Lower quality images will not age as well as higher quality images.

- High quality digital reproductions generally require better (and more expensive) equipment, more time to create, and have a higher storage footprint.
  - Higher quality images are more reusable
  - May be a more useful representation of the item for your users
  - Will remain useful for a longer time as image rendering technology improves and as technology evolves.
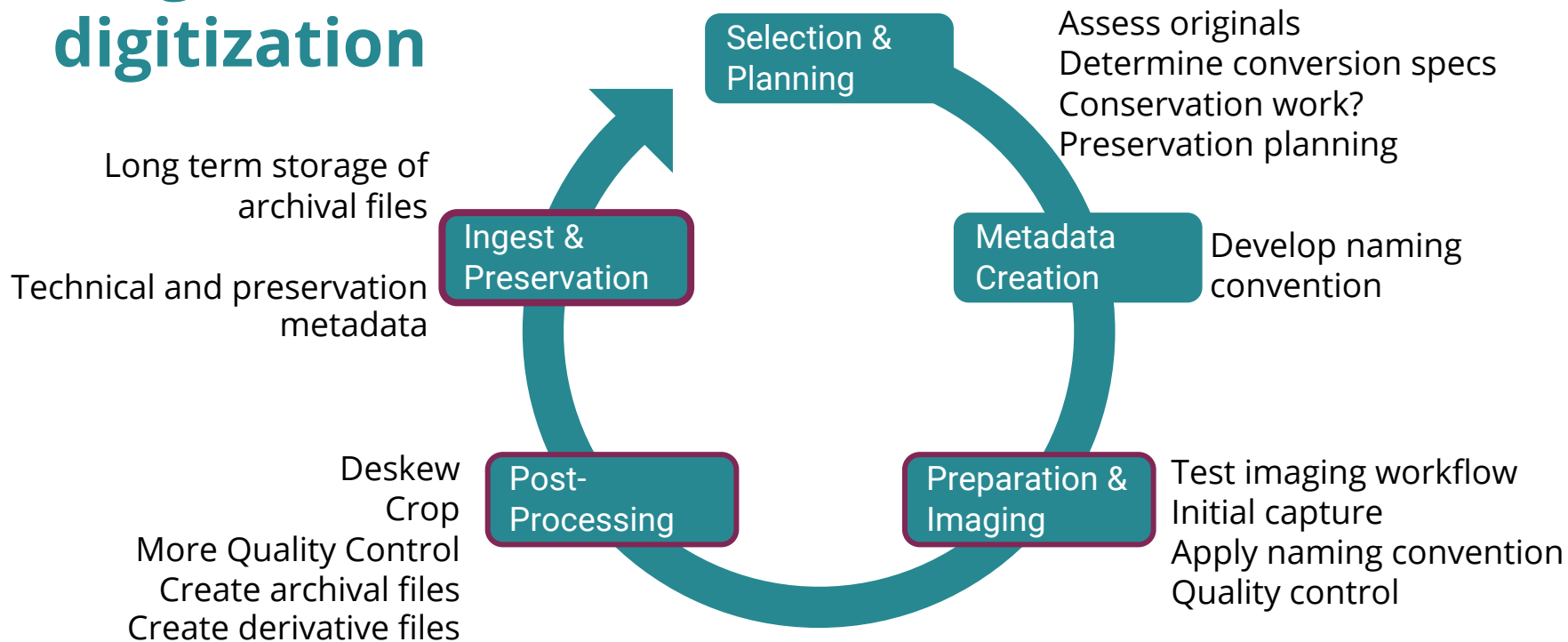
Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC LIBRARY SERVICE

Georgia's Virtual Library
GALILEO
An Initiative of the University System of Georgia

Imaging

Stages of digitization

Selection & Planning
- Assess originals
- Determine conversion specs
- Conservation work?
- Preservation planning

Metadata Creation
- Develop naming convention

Preparation & Imaging
- Test imaging workflow
- Initial capture
- Apply naming convention
- Quality control

Post-Processing
- Deskew
- Crop
- More Quality Control
- Create archival files
- Create derivative files

Ingest & Preservation
- Long term storage of archival files
- Technical and preservation metadata

What do you have already?

Evaluate equipment for

- Positioning
- Safe handling of materials
- Operator comfort and safety
- Stability
- Lighting
- Resolution and acuity
- Color fidelity
- File outputs
- Noise and compression artifacts

Digital Library of Georgia
Sharing Georgia's History & Culture Online

Georgia's Virtual Library
GALILEO
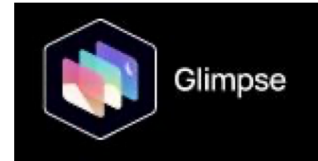An Initiative of the University System of Georgia

Evaluating imaging equipment for purchase

- Does it do what you need it to?
- What kind of files can it produce?
- Is it affordable?
- Is it easy to use?
- Can it be upgraded as technology improves?
- Can you talk to someone who has one?
- What is its highest optical resolution?
- Does it require proprietary software to use?
- Is it durable?
- Does it require a service or maintenance contract?

Image Editing Software

Minimum requirements:
- Deskew
- Crop
- Resize
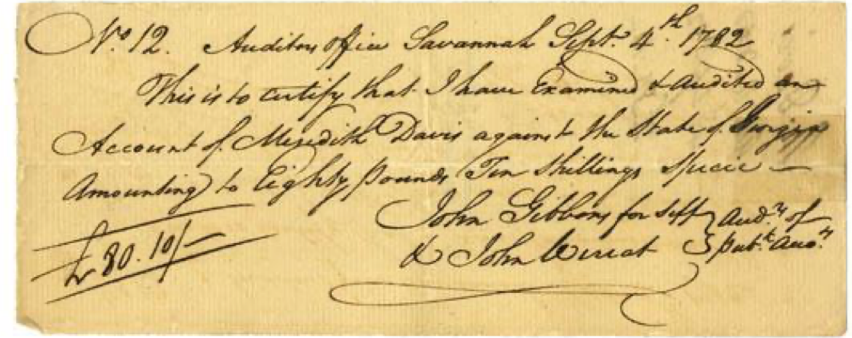- Save files to your chosen formats

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC LIBRARY SERVICE

Georgia's Virtual Library
GALILEO®
An Initiative of the University System of Georgia

Imaging for cultural heritage digitization

- Goal is to create an accurate representation of an object, not digital restoration.

- Safety first! Observe all materials handling guidelines.

- Assess materials to determine what you can safely capture with available equipment. There may be things that are too risky to scan.

  - Very brittle paper
  - Oversized items
  - Bound volumes

- Make sure donors/stakeholders are informed about any risks to items.

Hargrett Rare Book and Manuscript Library

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC LIBRARY SERVICE

Georgia's Virtual Library
GALILEO
An Initiative of the University System of Georgia

Recommended flatbed scanner settings

- For reflective items

  - 400 ppi
  - 24 bit RGB color
  - May need descreen for halftones
  - Auto-exposure off
  - Saved as uncompressed, single page TIFF (.tif)

- For transmissive items (film or slides)

  - 2400 ppi for 35 mm (or highest *optical* resolution)
  - 24 bit color or 8 bit gray (b/w)
  - Auto-exposure may be necessary.
  - Use thumbnail setting, but be careful of auto-crop.
  - Irregular sizes or dense negatives may need manual scan area selection.
  - Save as uncompressed, single page TIFF (.tif)

Check quality as you scan

Verify the quality of your images as you scan!

- Does the scan look like the object?
  - Is the white point set correctly?
  - Turn auto exposure off for historical materials.

- Has the entire object been captured?

- Is there light bleed at the edges of the image?
  - Grayed out areas where light enters around scanner lid.
  - If you can't close the lid, use a light proof drape.

Post-capture editing and quality control

Straightening and cropping

Open images in editor such as Photoshop

- Assess image quality at 100% size.
  - Does it look like the original?
  - Are all areas of the original visible?
  - Is text legible?
  - Is it free from dust, lint, and scanner artifacts (streaks)?
- Make sure they are oriented for easy reading.
- If images are crooked, use rotate feature to straighten them. Be careful not to distort image.
- Crop image leaving small border around the edge.
- Verify file name

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC
LIBRARY SERVICE

Georgia's Virtual Library
GALILEO®
An Initiative of the University System of Georgia

Creating derivative files for presentation

- What kinds of files you will need for online presentation?
  - Screen resolution JPEG (.jpg) files for online display (72 or 96 ppi)?
  - Files suitable for printing (at least 300 ppi)?
  - Multi-page PDF files?

- Create derivative files from the finalized archival tifs.
  - Individually as you review and crop archival files.
  - In batches using Photoshop or other editing software.

Online Delivery with the Digital Library of Georgia

Services include:

- Hosting of access or archival files
- Metadata funneling to DPLA
- Metadata hosting
- Metadata harvesting/ingest
- https://dlg.usg.edu/participate/partner-services

Have a project in mind? Check out DLG's Subgranting Program:

https://sites.google.com/view/DLG-Docs/dlg-documentation/dlg-subgranting-program

https://dlg.usg.edu

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC LIBRARY SERVICE

Georgia's Virtual Library
GALILEO
An Initiative of the University System of Georgia

Digital Preservation

Don't just think about it at the end!

Be sure to include preservation planning in your overall project plan:

- What files will you archive?
  - Which version?
  - What filetype?
- Where will you store them?
  - Local drives?
  - Cloud storage?
  - Backups/redundancy/geographic distribution?
  - Who has access?
- How will you maintain them over time?
  - What metadata do you need to preserve?
  - How to maintain linkages with descriptive metadata

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC LIBRARY SERVICE

Georgia's Virtual Library
GALILEO®
An Initiative of the University System of Georgia

Best practices for archival file formats

- Choose formats based on open, published standards.

- Use file formats suggested by LOC's Recommended Formats Statement for digitization projects to create archival files. http://www.loc.gov/preservation/resources/rfs/TOC.html

- LOC Sustainability of Digital Formats site https://www.loc.gov/preservation/digital/formats/

File fixity and data integrity

- Establish fixity using checksums

- Checksums are a representation of the actual bits of data in a file. If the data changes, the checksum changes.

- Using checksums allows you to detect changes to a file over time

- Checksums are an integral part of the LC BagIt specification.

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC
LIBRARY SERVICE

Georgia's Virtual Library
GALILEO®
An Initiative of the University System of Georgia

File packaging with the BagIt specification

- Packages files for transfer and storage

- Allows repositories to check completeness and fixity

- Platform independent

- Widely implemented

- Free tools in many different languages including Python, Java, and Perl

Safe storage practices

Safeguard your files!

Keep *at least two copies in different places* of your finalized archival files and the metadata required to understand what they are and how to render them in at least two different places, preferably on two different kinds of media.

- Local drive of workstation
- External drive
- Cloud storage (*know provider's policies on backups, downtime, and retrieval*)
- File server
- Avoid optical media such as CD or DVD

DLG archival storage services

- GALILEO partners may receive 30GB archival storage and associated production storage space.
- Non-GALILEO partners may receive of 20 GB archival storage storage and associated production storage space.
- Additional storage available to partners on a cost recovery basis.
- DLG Subgrant proposals may include archival storage services.
- Archival storage includes redundant, monitored storage in our designated preservation repository.

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC
LIBRARY SERVICE

Georgia's Virtual Library
GALILEO®
An Initiative of the University System of Georgia

More information

- An Introduction to Digital Preservation, Elizabeth La Beaud (Mississippi Digital Library). (DLG/HomePLACE sponsored webinar recorded 2019-03-28) https://drive.google.com/open?id=1ikXGTa6R_U8jdd6TV_RwBD8hhGXl5Z5T

- Community Owned digital Preservation Tool Registry. https://coptr.digipres.org/Main_Page

- Digital Preservation at the Library of Congress. https://www.loc.gov/preservation/digital/

- Digital Library of Georgia Presentations and Workshops. https://sites.google.com/view/dlg-docs/dlg-documentation/presentations-and-workshops

References

BagIt Python. (2019, April 04). Retrieved July 10, 2019 from
    https://github.com/LibraryOfCongress/bagit-python

Bagger. Retrieved July 10, 2019 from https://github.com/LibraryOfCongress/bagger

File Information Tool Set (FITS). Retrieved July 10, 2019 from
    https://projects.iq.harvard.edu/fits/home

Kunze, J., Littman, J., Madden, E., Scancella, J., & Adams, C. (2018, September 17). *The BagIt*
    file packaging format (V1.0) [IETF Internet-Draft]. Retrieved from
    https://tools.ietf.org/html/draft-kunze-bagit-17

Digital Library of Georgia
Sharing Georgia's History & Culture Online

GEORGIA PUBLIC LIBRARY SERVICE

Georgia's Virtual Library
GALILEO
An Initiative of the University System of Georgia

References

National Digital Stewardship Alliance. *Levels of digital preservation V2.0 matrix.* (n.d.). Retrieved June 11, 2020, from https://ndsa.org/publications/levels-of-digital-preservation/

Questions?

Angela Stanley, Director, Georgia HomePLACE
Georgia Public Library Service
astanley@georgialibraries.org

Mary Willoughby, Digital Conversion and Curation Librarian
Digital Library of Georgia
smirk@uga.edu