

Digitizing Materials at Small Archival Repositories



SARAH LEU

Lead Project Surveyor

Hidden Collections Initiative for Pennsylvania

Small Archival Repositories

Historical Society of Pennsylvania

sleu@hsp.org

<http://www.hsp.org/hcipsar>

Overview

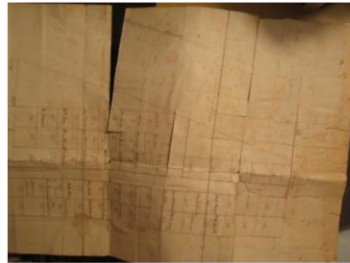


1. Why digitize materials?
2. Digitization process
3. Programs for sharing digitized materials
4. Additional resources

Why Digitize Materials?



PRESERVATION VS. ACCESS



Why Digitize Materials?



- **Preservation**

- Less handling of the original = reduced potential for damage
- Longevity- but only if you are prepared to migrate digitized materials from one generation of technology to the next

- **Access**

- Provides access to researchers unable to visit repository
- Improves access for oversized or fragile materials
- Creates opportunities for online exhibitions and use of images in promotional materials
- Increases overall exposure of collections and repository

Digitization Process

- 1. ORGANIZE**
- 2. CONSIDERATIONS**
- 3. SELECTION**
- 4. COPYRIGHT**
- 5. SCANNING**
- 6. METADATA**

Digitization Process: Organize



Before digitizing: Arrange and describe your collections!



Image rights: Fireman's Hall Museum

You don't have to organize every single collection before digitizing, but whatever materials you choose to digitize should be part of a collection that is already organized. The beginning of a digitization project is a good time evaluate collection conditions and rehouse and arrange materials.

Digitization Process: Organize



Before digitizing: Arrange and describe your collections!

- Helpful for:
 - Naming digital files consistently
 - Locating materials in the future
 - Identifying missing items
 - Returning misplaced items



Image rights: Fireman's Hall Museum

If someone wants to see other materials from the same collection as the digitized item, you can locate the collection easily, and give them a finding aid to read about the other materials in the collection and then locate any other materials they want to see.

If things are just scattered in a box without labeled folders, you might not know if material is missing, if you have a written record or a folder labeled saying it was there at some point, but it is no longer there, you know something is missing.

If someone removed material to be digitized, but then got side-tracked with other things and then that person no longer works for the organization and someone else finds the materials, they will know where they go.

Digitization Process: Considerations



Things to think about while planning a digitization project.



Digitization Process: Considerations



- **Is this type of project a priority?**
 - Evaluate user needs, repository goals, and collections
- **What would the scope of the project be?**
 - Define goals and requirements of the project; duration
- **Is this feasible with our budget and available space?**
 - Costs of equipment, software, training, staffing
 - Space for equipment and storage space for digitized materials
- **Is it sustainable?**
 - Long-term management and maintenance of images and metadata

Digitization Process: Selection



What materials will be digitized?

Develop selection criteria.

You can have different criteria for different projects, can weigh criteria differently, but selected materials should always fit with your repository's mission and the goals of the digitization project (access, preservation, new audiences, promotion).

Digitization Process: Selection



Selection Criteria

- Use: Is there a high demand for these materials?
- Value: Are materials especially significant to your audience? Do materials have a unique value or cover subjects not well represented elsewhere?
- Condition: Are the materials at risk because of poor housing, high use, or physical deterioration? Would digitization worsen condition/cause damage? Is the collection arranged and described already?

Digitization Process: Selection



Selection Criteria (continued)

- Format: Would a digitized version be easier to access than original format? Do you have equipment to digitize this format?
- Are the materials appealing to the funder or institution?
- Will there be enough institutional support to maintain the digitized format?
- Can the materials be captured adequately enough to be a surrogate in place of the original (without inflicting damage to original)?

Digitization Process: Copyright



Something else to consider when selecting materials...



Digitization Process: Copyright



Copyright and Permissions

- Before you digitize (or put materials online), check your deed of gift form to see if you have the right to reformat, copy, and/or distribute the materials.
 - Having physical ownership does not mean that you also control the intellectual property rights of the item(s).
 - Unless materials are in the public domain, intellectual property rights must be granted/transferred in writing.
 - Transfer of copyright can only be done by the creator or person/organization that holds the copyright- not always the same person who has physical ownership at time of acquisition.

If you do not have a transfer of intellectual property rights statement or an outline for permissions regarding format transfer on your deed of gift or donor form, add it now.

Digitization Process: Copyright



Copyright and Permissions

- Your repository can digitize copyrighted materials for preservation purposes, but these images must only be accessed on-site, not placed online, which would be considered publishing the image.

If you want to place images online, you will need to determine who owns the intellectual property rights such as copyright, trademark, patent, publicity rights, performance rights etc.

Digitization Process: Copyright



Copyright and Permissions

- If copyright has been legally transferred to your institution or the materials are in the public domain, you can proceed without risk.
 - To determine what is in the public domain or what might be in copyright, see the public domain chart by Peter B. Hirtle of Cornell University: <http://copyright.cornell.edu/resources/publicdomain.cfm>.
 - If materials are in the public domain, have a disclaimer that states that the item is in the public domain, that there are no known use restrictions, and to contact the repository for a high-resolution copy. If your institution holds the copyright, include a disclaimer that states the item is in copyright and to contact your repository for information about permissions and reproductions.

Never published, never registered works will be in the public domain after the life of author plus 70 yrs, so in 2015, never published or registered works of authors who died before 1945 are in the public domain. Published works before 1923 are in public domain.

You should also think about a preferred citation method for your materials in the event someone wishes to use them. See Wesleyan University's Citing Archival Resources page for some ideas:
<http://www.wesleyan.edu/libr/schome/discover/citingresources.html>.

Also, think about how to handle someone asking for a high resolution version. Will you send it to them? In what format? Will you charge? How much?

Digitization Process: Copyright



Copyright and Permissions

- If the item is in copyright and the intellectual property rights have not been transferred to your institution, contact the copyright holder to request permission to publish the image(s) online.
 - If permission is obtained, include a disclaimer with the image. State that the item has been made available with permission of the copyright holder and has been provided on your site for educational purposes, and that commercial use is prohibited without permission. Note that the researcher should contact the repository for information about permissions and reproductions.

Digitization Process: Copyright



Copyright and Permissions

- If an item is believed to be in copyright, but you cannot identify or locate the copyright holder (orphan work), you should document your repository's attempts to identify or locate the copyright holder as proof of due diligence.
- If you still cannot locate or identify the copyright holder and choose to publish the image(s) online, include a disclaimer noting that the item may be protected under copyright law, that the user is responsible for all issues of copyright and to contact the repository for information regarding permissions and reproductions.
- Placing an image online in this situation is done at your own risk, and you should be prepared to remove the image(s) if challenged.
 - Consider creating a policy for taking images down from your site.

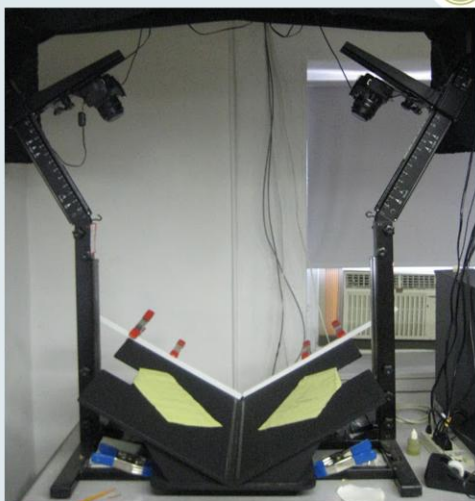
Digitization Process: Copyright



Copyright and Permissions

- Ultimately, copyright is about risk management. Is it likely someone will protest your digitization of the materials? Do you have the resources to deal with the consequences?
- Use common sense. Are the materials older and has the creator likely been dead for at least 70 years or are they more recent materials, created by someone of prominence, or likely to get a lot of use?

Digitization Process: Scanning



Equipment
Settings
Storage
Issues

Digitization Process: Scanning

- Use a scanner with low platen temperature
 - Flatbed for standard sized documents, overhead or digital camera for oversized documents and bound volumes
- 400 PPI, 24-bit color (RGB), TIFF (.tif) format
- Keep a master unedited copy (.tif)
 - Plus smaller **access** copies (.jpeg, .pdf)
- Sufficient and safe storage space
 - At least two master copies (unedited) stored in different buildings
- Establish file naming conventions
 - CollectionNumber_BoxNumber_FolderNumber_ItemName (e.g. 012_03_07_ajohnsonportrait)

Different formats have different scanning standards, but overall 300 or 400 PPI should suit your purposes.

Access copies are the ones you put online and can be edited. Always retain two unedited master copies in the .tif format. You may also wish to save an edited high resolution copy, but do not replace your master copies with this.

File names should be a unique identifier for item.

Use leading zeros before numbers to make sure file names are the same length and stay in file name order when sorted.

Use only alpha-numeric characters for both files and folders. Exceptions are dashes (-) and underscores (_).

Use a valid file extension, which is usually has three and sometimes four characters (i.e. .tif .pdf .html, etc.).

Do not use spaces in file/folder names – use dashes or underscores instead.

When dates are used, choose a standard format.

Be brief! Try to keep file names at 25 characters or less, if possible. Long file names may prohibit portability.

File names should contain necessary descriptive information, independent of their storage location.

If you need to denote that the file is a particular version of an image, include a version numeral.

Digitization Process: Scanning



Potential Issues

- Bound documents can be too large or fragile to scan without specialized equipment
- Audio and video media (audiocassettes, reel-to-reel, VHS) can be difficult to digitize due to technological obsolescence
 - Similar issues also apply to computer storage units
- Photographic materials (slides, negatives, old photographs, glass plates, and lantern slides) can require use of specialized equipment
- Digital versions sometimes result in a loss of detail or color
- What to do with the originals after scanning?

Formats of the materials can be of interest to researchers too; they represent advances or limits of technology from the time they were created. Repositories should consider creating a policy for what to do with originals after digitization.

Digitization Process: Metadata




Metadata is “data about data”

Metadata is used to describe, locate,
retrieve, and share your materials.

The types of data you record will really be up to you. It may vary by collection and ultimately relies on your institutional needs. Keep in mind that you may not need to re-invent the wheel with your metadata, you might already have descriptive metadata in a database, spreadsheet or finding aid.

Digitization Process: Metadata



Types of Metadata	Used to	Example
Descriptive	Describe the who, what , where, and when	Subject, image summary/title, date
Structural	Describe the physical nature of the item	Number of pages, file format, size
Administrative	Manage the item	Rights & reproduction information, date of digitization
Preservation	Protect the item from deterioration and destruction	Document preservation actions, ensure authenticity over time

Chart from: Missouri Historical Records Advisory Board Digitization Guidelines for Small Historical Institutions and Repositories. Last updated 2013.

http://s1.sos.mo.gov/CMSImages/Archives/MHRAB_Digitization_Guidelines.pdf .

Digitization Process: Metadata

Basic Metadata Elements

- Collection
- Box number
- Folder number
- Image identifier
- Folder title
- Summary or item title
- Date
- Location
- Subjects or other notes

Metadata does not have to be complex.

Example of basic metadata elements from: Missouri Historical Records Advisory Board Digitization Guidelines for Small Historical Institutions and Repositories. Last updated 2013.

http://s1.sos.mo.gov/CMSImages/Archives/MHRAB_Digitization_Guidelines.pdf.

Digitization Process: Metadata



Using **standardized** metadata is a must if you want to ensure your materials are **findable** by researchers and that your descriptions are **shareable** with other computer systems

Metadata Standards	Used to	Examples
Value Standards	Determine terms used to name people and subjects represented	Library of Congress Subject Headings and Name Authority Files (LCSH, NAF)
Content Standards	Determine what kinds of information you use in the description	Describing Archives A Content Standard (DACS)
Format Standards	Determine how the metadata should be structured/encoded	Dublin Core Metadata Initiative (DCMI)

The importance of having a standard to name digital files was already mentioned. Here are some other types of metadata standards.

Programs for Sharing Digitized Materials



1. OMEKA
2. VIEWSHARE
3. PASTPERFECT



Free for basic (500 MB storage), low-cost for enhanced features and additional storage.

They host the website for you (no special server requirements).

Good for online exhibitions with pictures and text.

<http://omeka.net>



The screenshot shows a website titled 'THE LAND OF PENN & PLENTY' with the tagline 'Bringing History to the Table'. It features a navigation bar with links for About, Exhibits, Events, Articles, and Items. The main content area is titled 'PENNSYLVANIA HISTORICAL MARKERS' and includes a search bar, a list of markers (Birthplace of Commercial Ice Cream Production, Hotel Fauchere, Arnold Naeffels, Loretta Langstroff), and a detailed article about the 'Birthplace of Commercial Ice Cream Production' featuring images of a historical marker, a portrait of Jacob Smysier, and a vintage ice cream advertisement.

Pennsylvania Foodways website. For info on using Omeka with archives, visit: <http://info.omeka.net/omeka-net-help/use-case-archivists/>. For general info, visit: <http://info.omeka.net/about/>. Omeka allows you to track website visits with a Google Analytics plugin. Omeka uses Dublin Core as a format standard. You can upgrade at any time. See levels of enhanced storage options below.

Omeka.net basic site: Free, 500 MB storage, 1 site, 14 plugins, 4 themes

Omeka Plus site: \$49/year, 1 GB, 2 sites, 19 plugins, all themes

Omeka Silver site: \$99, 2GB, 5 sites, 26 plugins, all themes

Omeka Gold site: \$299, 5GB, 10 sites, all plugins and themes

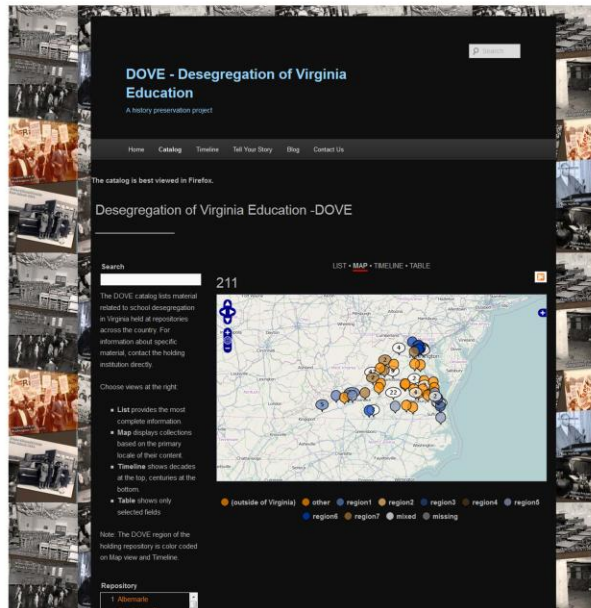
Omeka Platinum site: \$999, 25GB, unlimited sites, plugins, and themes




Free.

Good for analyzing data and making different types of visualizations (map, timeline, chart, list) that you can embed on your website.

<http://viewshare.org>



Hosted by Library of Congress.



\$870 + \$670 for
MultiMedia and
PastPerfect Online
Setup

They will host a site
for you for an
annual fee of \$440
(up to 10,000
catalog records) .

Integrated with
PastPerfect catalogs.


<http://museumsoftware.com/>

African American Museum in Philadelphia
Honoring our legacy Embracing our promise
www.aampmuseum.org


[Home](#)
[Click & Search](#)
[Keyword Search](#)
[Advanced Search](#)
[Random Images](#)
[Search Results](#)
[Help](#)

Random Images


[More Random Images](#)




1986.001




1986.001




2010.001




1986.001




2010.001




1986.001



1986.001



2010.001



1986.001

[More Random Images](#)

Copyright © 2012 The African American Museum in Philadelphia. Reproduction without permission is prohibited.
This is an JPL® Online Catalog.

PastPerfect's HQ is in Exton, PA. They provide excellent customer support and training. PastPerfect provides a watermark and resizing feature, works on mobile phones, and your materials will be indexed by Google, making them accessible through a Google search. There is a 20% discount on PastPerfect if your institution is a member of the American Association for State and Local History (\$115 for basic institutional membership). Altogether at regular price, PastPerfect with everything you need to put images online is \$1980, with the discount the price is \$1584. Keep in mind the fee for PastPerfect hosting your site (\$440) will be charged annually.

One More Time...



**DON'T FORGET TO MAINTAIN YOUR
DIGITIZED FILES AND METADATA!!**

Additional Resources



Additional Resources



Digitization Guidelines and Programs

- Missouri Historical Records Advisory Board Digitization Guidelines for Small Historical Institutions and Repositories
http://s1.sos.mo.gov/CMSImages/Archives/MHRAB_Digitization_Guidelines.pdf
- Library of Congress Preservation Digital Reformatting Program site:
<http://www.loc.gov/preservation/about/prd/presdig/index.html>

Selection Criteria

- University of California Libraries selection criteria
<http://libraries.universityofcalifornia.edu/content/university-california-selection-criteria-digitization-pag>

Copyright

- Public domain chart by Peter B. Hirtle of Cornell University
<http://copyright.cornell.edu/resources/publicdomain.cfm>

Additional Resources



Format Metadata Standard

- Dublin Core (15 core metadata elements):
<http://dublincore.org>

Content Metadata Standard

- Describing Archives: A Content Standard (DACS):
<http://files.archivists.org/pubs/DACS2E-2013.pdf>

Standardized Vocabularies

- Library of Congress Subject Headings and Name Authority File: <http://id.loc.gov>
- Getty Vocabularies Art & Architecture Thesaurus:
<http://www.getty.edu/research/tools/vocabularies>

Additional Resources



Programs

- Omeka: <http://info.omeka.net/about/>
 - Using Omeka for archives: <http://info.omeka.net/omeka-net-help/use-case-archivists>
- Viewshare: <http://viewshare.org>
- PastPerfect: <http://museumsoftware.com/>
 - HCI-PSAR guide on how to optimize PastPerfect for use with archival collections: <http://hsp.org/hcipsar/resources>

Digitizing Materials at Small Archival Repositories



SARAH LEU

Lead Project Surveyor

Hidden Collections Initiative for Pennsylvania

Small Archival Repositories

Historical Society of Pennsylvania

sleu@hsp.org

<http://www.hsp.org/hcipsar>