DHPS | NY

DOCUMENTARY HERITAGE & PRESERVATION SERVICES FOR NEW YORK

Introduction to Digitization: Webinar

Tamara Talansky, Preservation and Imaging Project Coordinator
Conservation Center for Art and Historic Artifacts
December 8, 2016

Documentary Heritage and Preservation Services for New York is a five year initiative to deliver collections-related training, preservation surveys, archival assessments, and other services to the historical records community in New York.

New York State Library

New York State Archives

DHPS | NY

Documentary Heritage & Preservation Services

for New York

Website: dhpsny.org

Email: info@dhpsny.org

The University of the State of New York — The State Education Department — 1784

Conservation Center

for art & historic artifacts

Introduction to Digitization

Digitization vs. Digital Preservation

- Digitization:
  - Definition from the Federal Agencies Digitization Guidelines Initiative (FADGI):
  - "Digitization is the process of recording an analog signal in digital form…commonly for increased access or for preservation purposes."

Digitization vs. Digital Preservation

- Digital Preservation:
  - Definition from the American Library Association (ALA):
  - "Digital preservation combines policies, strategies and actions to ensure access to reformatted and born-digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time."

Differences

- Digitization:
  - Scanning
  - Actual transfer of information

- Digital Preservation:
  - Long-term commitment
  - Policy development
  - Dedication to the content

Digital Preservation Polices

- Create long-term repository
- Needs and methodologies
- Guarantees continuing access

Digital Preservation Policies

- Purpose
- Goals
- Challenges
- Best practices
- Stakeholders
- Content types
- Tool for funding collections care
- Grant applications

A Plan is a Must

- Why are you digitizing?
- Who is the audience?
- Who else needs to be involved?
- What should you digitize?
- Are you scanning for preservation purposes?
- How about copyright issues?
- What about hardware, software and best practices?
- In-house or Outsourced?
- How do you manage and deliver the files?

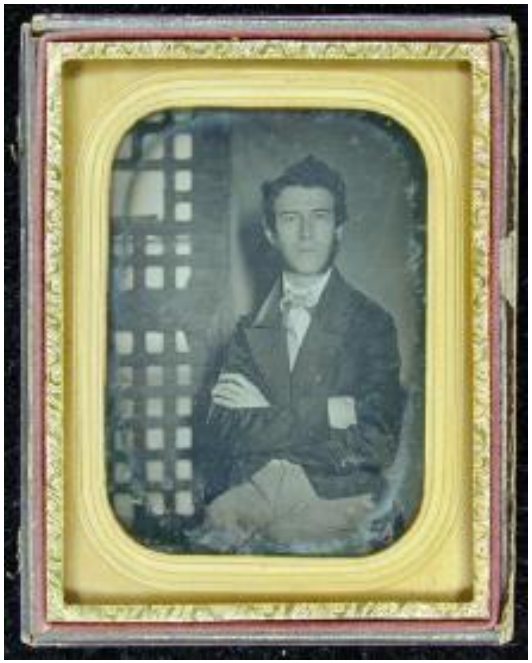
Why digitize?

- Access / Outreach
- Preservation
- Storage Needs

Credit: http://hcl.harvard.edu/collections/digital_collections/digitization_program.cfm

Accessibility

- Access all types of collections
- Broader audiences
- Easily search online
- Outreach to scholarly communities and social media



Credit: Chester County Historical Society

Preservation

- Digital surrogates protect originals
- Provides access while keeping originals secure

Digitization is not preservation, but…

- Reduces handling of original
- Preferred for paper materials
- Widely embraced by scholars

- Drawbacks:
  - Media obsolescence concerns
    - Technical dependency
    - Media deterioration
  - Costs

Credit: Fisk University, Nashville, TN

Storage Constraints

- Lack of space
- Materials prone to damage
- Document vulnerability
- Related costs

Credit: http://shutterstock.com

Audience and Workflow

Who is Your Audience?

- Cultural organizations
- General users
- Students
- Researchers
- Genealogists
- Business community
- Universities

Knowing Your Audience

- Audience Determines:
  - Materials
  - Resolution
  - Equipment
  - Technical choices
- How will surrogates be used?
  - Publications
  - Printed
  - Web-based

Stakeholders

- In-house staff
- Project staff
- Tech support
- Partners outside of organization
- Volunteers
- Collaborations
- Potential funders
- General Users

Credit: Digital & Preservation Services, LYRASIS

Staff and Workflow

- What needs to be done:
  - Selection of materials
  - Scanning
  - Metadata
  - File management/backups
  - Quality Control
  - Website design and technical support
  - Project management

Credit: LYRASIS

Staff and Workflow

- Additional Skills
  - Conservation
  - Photographic skills
  - Database admin
  - Computer work

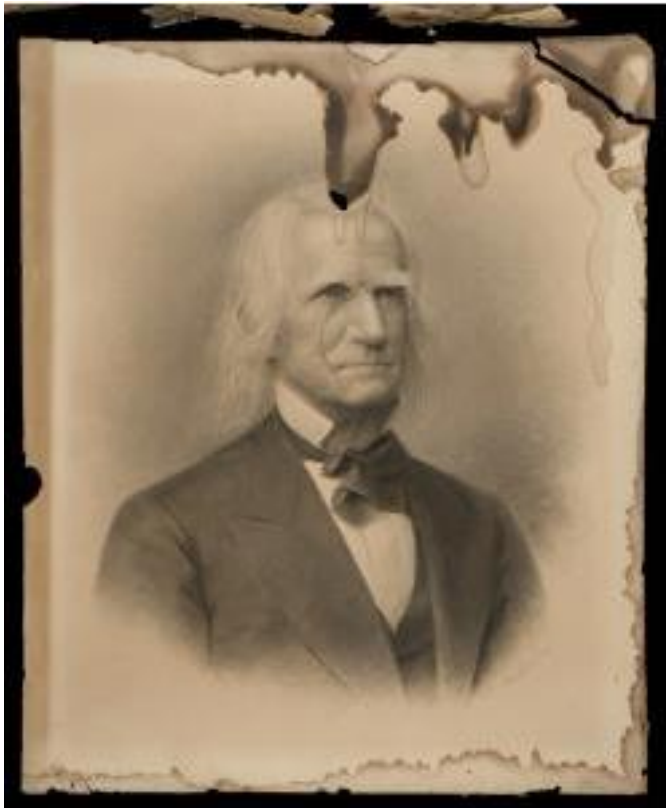
Project Management

- Setting timelines
- Determining approaches for accomplishing goals
- Developing budgets
- Communicating with participants / outside vendors
- Monitoring production
- Looking beyond project's end

Materials and Copyright

Materials Selection

- Condition
- Stabilizing treatment
- Content
- Uniqueness
- Use/Demand
- Accessibility
- Cultural significance
- Audience
- Mission



Sample Collections Matrix

Institution's Collections	Publications	Photographs	Manuscripts	XYZ Archives	Diaries / Scrapbooks	Oral Histories	Newspapers	Maps/Large Format flat	Periodicals
Variable	1-5 points								
Usage level/demand (H,M,L)				Low, 1					
Cultural significance (H,M,L)									
Accessibility (H,M,L)									
Uniqueness (H,M,L)				High, 5					
Intellectual Value (H,M,L)									
Rights Cleared (Y/N)									
Metadata (Y/N)									
Physical condition (F/G)				Fragile, 1					
Advances mission (Y/N)									
Geographic area (Y/N)									
Audience (Internal Staff, Researchers, General Public)									
Technological issues identified									
Research available for this collection (Y/N)									
Stabilizing Treatment (Y/N)				Yes					
Other									
TOTAL Points Value									

This chart could be used for multiple collections as you see it here. Or, you can adapt if for one Collection e.g. if a collection has diaries, photographs, scrapbooks, letters, manuscripts and oral history interviews…

Suggested points value 1= lowest and 5 = highest

H= High=5 points M= Medium=3 points L= Low=1 point

F= Fragile=1 point G= Good=5 points

Yes=5 points No=1 point

You will identify your own audiences and determine which audiences are considered primary for your digital projects—genealogists might be a primary audience for your institution whereas at a university faculty might be the primary audience and alumni the secondary audience. You decide as a group.

Staff and contractors (primary) depending on the project you might assign 5 points if you want to focus on this as a primary audience

R=Researchers (Secondary) depending on the project you might assign 5 points if you want to focus on this as your secondary audience

GP=General Public (Tertiary) depending on the project you might assign 5 points if you wanted to focus on a tertiary audience

Technology issues: Are there any special needs for this material? E.g., Zooming, streaming of content, other (Not to be ranked) You could also add notes and preservation concerns here.

Materials That May Be Excluded From the Digitization Process

Some items require more involved and expensive repair, or are so fragile that they cannot be easily stabilized. Some examples of materials that may be excluded from digitization are:

- Paper that is acidic, fragile, brittle, torn, missing pieces, sticky or stuck to something.
- Paper documents with three-dimensional objects (e.g., medals, seals, ribbons) attached.
- Paper with iron gall ink that has eaten into the page.
- Loose, flaking or friable media such as crayon, charcoal, chalk, or soft pencil.
- Books with severe leather deterioration (i.e., red rot) or missing pages.
- Letter copy books with very thin and acidic paper, such as carbon copy correspondence and some tracing paper drawings and plans.
- Photographs that are separating from the mount or support.
- Photographs that are curled, bent, creased, folded, wrinkled, cockled or cracked.
- Photographs and their mounts which are acidic, fragile, brittle, torn, missing pieces, sticky or stuck to one another.
- Deteriorated cellulose nitrate or acetate film negatives and positives.
- Scrolls or other non-traditional textual or image formats.

Credit: http://www.loc.gov/preservation/care

Legal Concerns

- Ownership vs. Copyright
  - Do you own it?
- Does donor hold rights?
- Revisit deed of gift
  - Modify to include rights or permissions

Credit: Robert W. Woodruff Library

Copyright Resources

- Lolly Gasaway, University of North Carolina http://www.unc.edu/~unclng/public-d.htm
- US Copyright Office www.copyright.gov
- Copyright Basics: Fair Use Checklist http://citl.indiana.edu/files/pdf/fair_use_checklist.pdf
- 10 Big Myths about copyright http://www.templetons.com/brad/copymyths.html
- Creative Commons http://creativecommons.org/about/licenses/
- National Initiative for a Networked Cultural Heritage http://www.ninch.org/copyright/
- Code of Best Practices in Fair use for Academic and Research Libraries http://www.centerforsocialmedia.org/libraries
- Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1495365 http://copyright.cornell.edu/
- LYRASIS Digital Toolbox: Rights Management https://www.lyrasis.org/LYRASIS%20Digital/Pages/Preservation%20Services/Resources%20and%20Publications/Digital%20Toolbox/Rights-Management.aspx

Credit: LYRASIS

Quality Control

Best Practices and Principles

- Must adhere to current industry standards
  - Selection
  - Acquisition
  - Upkeep
  - Distribution

Besser Principles

- Capture at the highest resolution appropriate to the informational content of the originals
- Capture at an appropriate level of quality to avoid rescanning and re-handling of the originals in the future--scan once
- Create and store a master file that can be used to produce derivative files and serve a variety of current and future user needs
- Use system components that are non-proprietary
- Use image file formats and compression techniques that conform to industry standards

Credit: Adapted from LYRASIS

Besser Principles

- Create backup copies of all files on a stable medium
- Create meaningful metadata for files or collections
- Store media in an appropriate environment
- Monitor and recopy data as necessary
- Outline a migration strategy for transferring data across generations of technology
- Anticipate and plan for future technological developments

Credit: Adapted from LYRASIS

International Federation of
Library Associations and Institutions

Guidelines for Planning the Digitization of Rare Book and Manuscript Collections

Written by the IFLA Rare Book and Special Collections Section

September 2014
Endorsed by the IFLA Professional Committee

http://www.ifla.org/files/assets/rare-books-and-manuscripts/rbms-guidelines/guidelines-for-planning-digitization.pdf

Best Practices Resources

- The Library of Congress Technical Standards for Digital Conversion
  https://memory.loc.gov/ammem/about/techStandards.pdf
- American Library Association, Minimum Digitization Capture Recommendations
  http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations
- U.S. National Archives and Records Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access
  http://www.archives.gov/preservation/technical/guidelines.pdf
- BCRs CDP Digital Imaging Best Practices Version 2
  http://sustainableheritagenetwork.org/system/files/atoms/file/bcrcdpImagingBP.pdf
- Open Archival Information System Reference Model (OAIS)
  http://www.oclc.org/research/publications/library/2000/lavoie-oais.html
- PREMIS Preservation Metadata Implementation Strategies
  http://www.loc.gov/standards/premis/
  http://www.loc.gov/standards/premis/v2/premis-2-0.pdf
- Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist
  https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf
- Federal Agencies Digitization Guidelines Initiative (FADGI): Draft of Technical Guidelines for Digitizing Cultural Heritage Materials
  http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image_Tech_Guidelines_2015-09-02_v4.pdf

Standards Organizations

- National and International Standards Organizations
  (from the Library of Congress): https://www.loc.gov/standards/

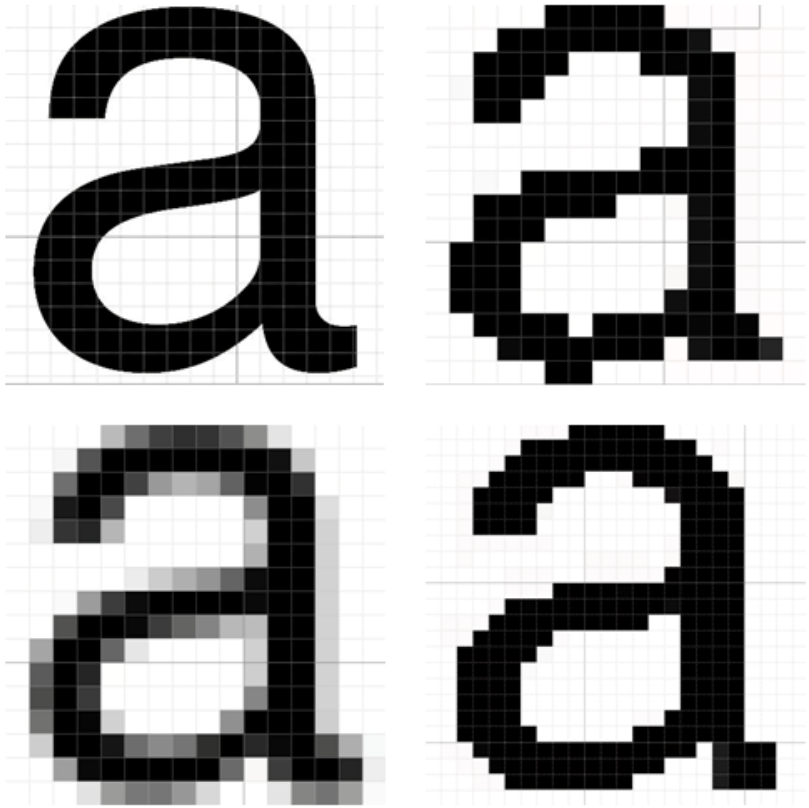
  - National Information Standards Organization (NISO)
    NISO is an ANSI-accredited organization that develops standards specifically for the library, information services, and publishing sectors.
    http://www.niso.org/

  - American National Standards Institute (ANSI)
    ANSI is the organization that facilitates development of American National Standards (ANSs) by establishing consensus among qualified groups.
    http://www.ansi.org/

  - International Organization for Standardization (ISO)
    ISO is the standards body that establishes standards for the international exchange of goods and services.
    http://www.iso.ch/

Digitization Basics

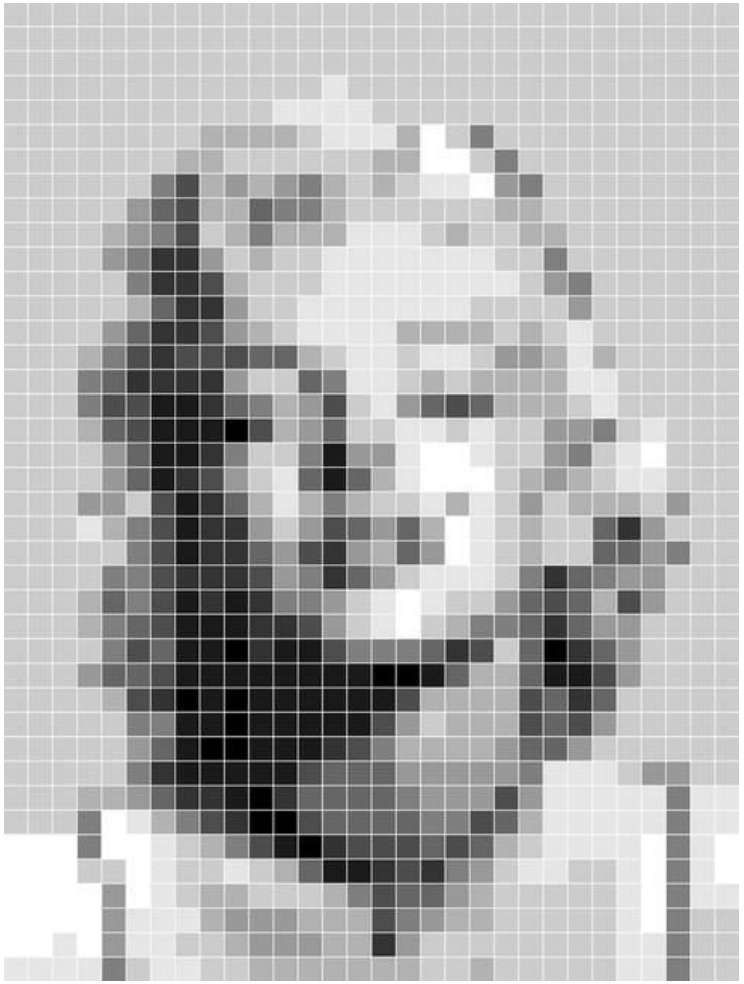
- Pixels
- Bit Depth
- Resolution



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

Pixels

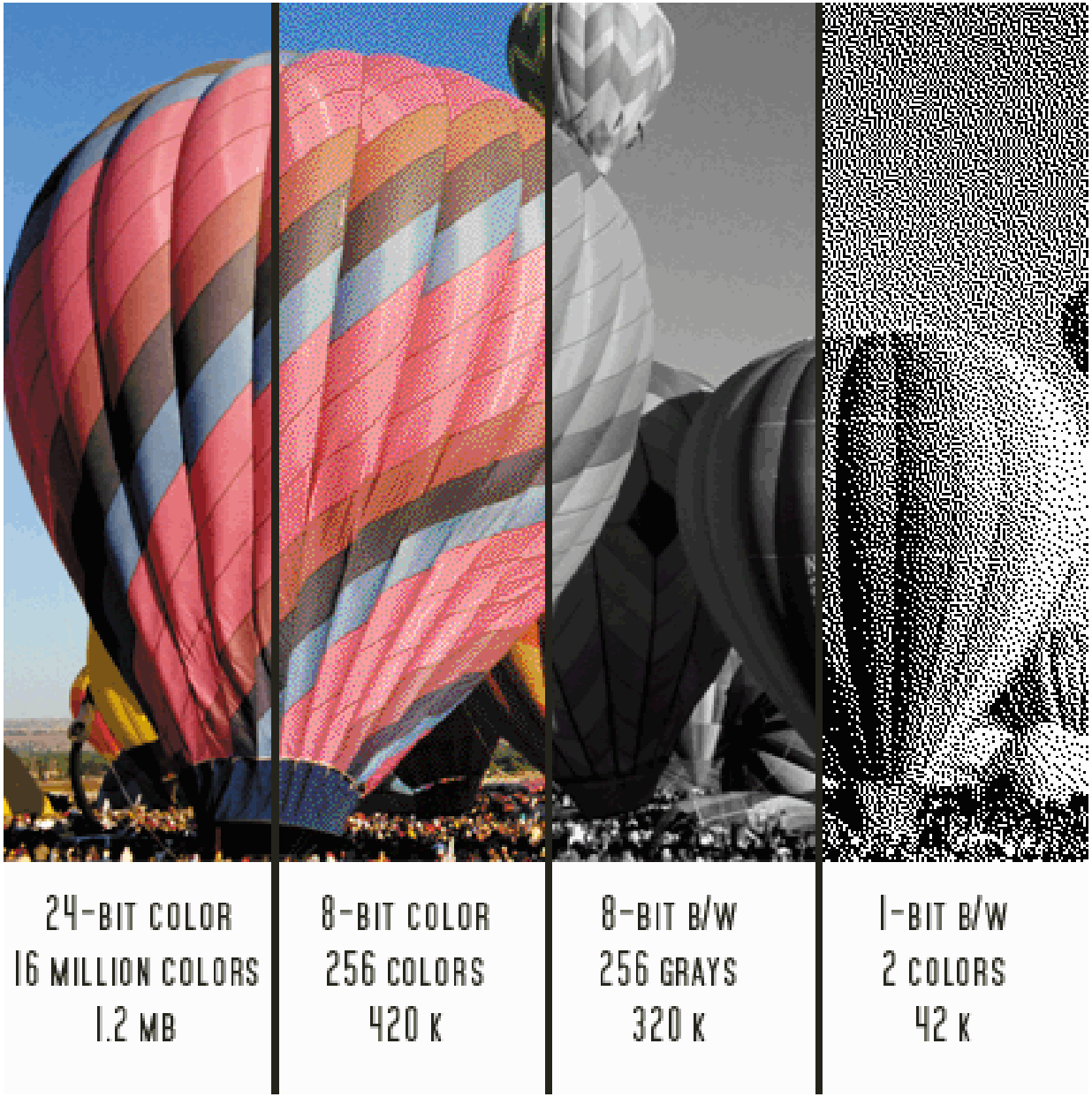
- Picture Elements = Pixels
- Tiny dots make up image
- Dimensions of a photo
  - 800 x 600
  - 1520 x 1280



Credit: http://booksat.scarlettrugers.com/

Bit Depth

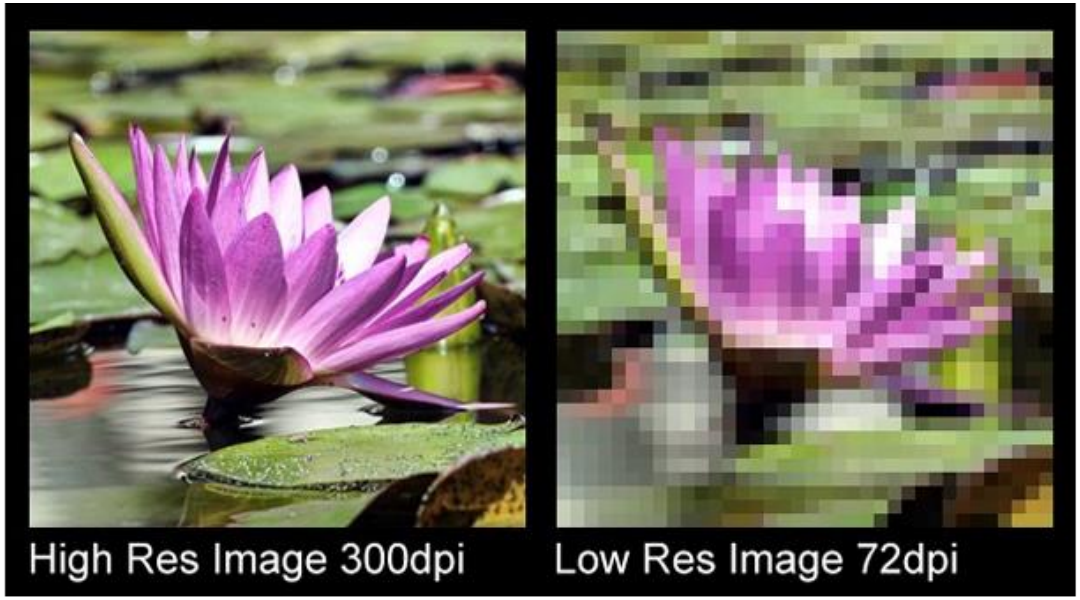
- Number of bits of data per pixel
- The higher the bit depth, the more info the pixel contains = better quality
  - Black and white =1 bit per pixel (2 colors)
  - Greyscale = 8 bits per pixel (256 shades of grey)
  - Color = 24 bits per pixel (16.7 million color tones)



Credit: veterangames.wordpress.com/2014/11/

Resolution

- Determined by number of pixels
- Dots-per-inch (DPI) / Pixels-per-inch (PPI)
- The higher the resolution, the finer the detail



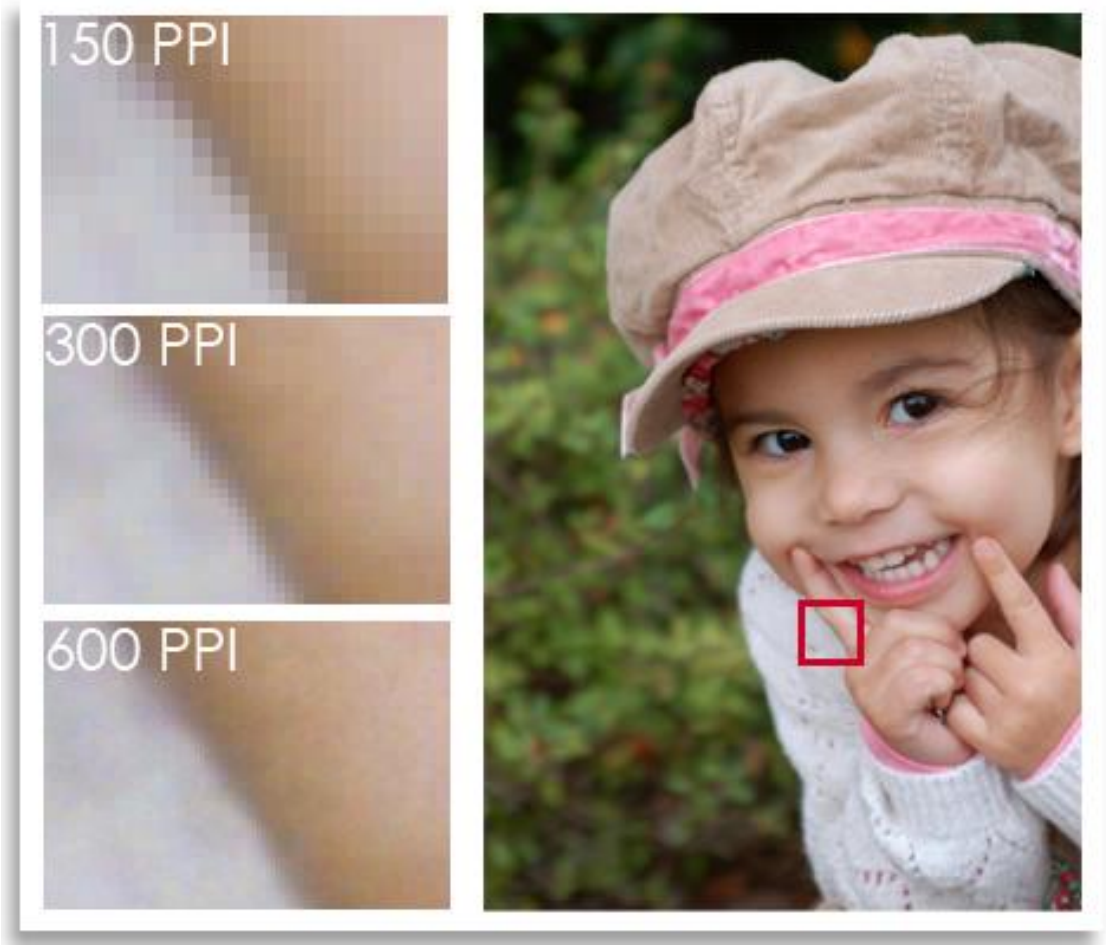
Credit:. http://ccideas.com/

Resolution

- Optimum settings
  - Common snapshots and enlargements
    - 300-400 DPI
  - Small prints/slides
    - 1400-1500 DPI
  - Negatives
    - 1500-2000 DPI

- Increasing DPI will always increase file size

Resolution



Credit: http://www.nashtucky.com/photoscanning.html

Resources

- Minimum Digitization Capture Recommendations http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations#photographic_processes

- Cornell Digital Imaging Tutorial http://www.library.cornell.edu/preservation/tutorial/contents.html

Most Common File Formats

- TIFF, JPEG, and PDF files

Credit: http://uofgts.com/PS-P2Site/settings.html

Image Compression

- Lossless
  - Preserves original data
  - Represents image without compromising accuracy.
  - Common in TIFF images.
- Lossy
  - Discards most of the data
  - Quality is lost.
  - Common in JPEG images.

Credit: https://ucscsgdfrancisdeoliveira.wordpress.com/author/francisdeoliveira/page/2/

TIFF vs. JPEG

- Tiffs are uncompressed
  - Retains maximum amount of detail
  - Can use LZW compression and maintain all data

- Jpegs are compressed
  - Loses data captured by the scanner

JPEG Quality

Maximum quality

Lowest Quality

Credit: http://photo.stackexchange.com/

TIFF: Tagged Image File Format

Advantages

- Long-term archiving
- Lossless compression
- Good for printing/editing
- High quality
- Up to a bit depth of 64
- Metadata tagging

Disadvantages

- Very large file
- Long transfer time
- Takes up space
- Slow loading time

JPEG: Joint Photographic Experts Group

Advantages

- Most widely accepted
- Standard for web
- Takes up little space
- Fast loading time
- Up to a bit depth of 24

Disadvantages

- Low quality
- High lossy compression
- Not good for editing/printing
- Should not use to archive

JPEG Compression

Credit: http://www.makeuseof.com/tag/how-does-file-compression-work/

PDF: Portable Document Format

- Captures printed intent
- Can be used by anyone
- Adobe Reader is free
- PDF's are compressed
- View and print at high resolution
- Easy to upload
- PDF/A for long-term archiving

Bottom Line…

- Always keep a TIFF Master!
  - Make TIFF copy for editing
  - Make JPEG from each master
  - Do not edit the master

Metadata

- Data that describes data
- Tells you something about the image
- Helps identify files
- Search by keyword or tag

Types of Metadata

- Descriptive
- Structural
- Administrative
  - Rights Management
  - Preservation

Metadata Best Practices resource:

CDP Dublin Core Metadata Best Practices 2.1.1

http://sustainableheritagenetwork.org/system/files/atoms/file/CDPDublinCoreBPs_0.pdf

Descriptive Metadata

- Describes object
- Keyword, author, title, etc.
- Helps create finding aids

Structural Metadata

- Relationships between objects
- Indicates structural division
  - Order in a chapter
  - Diary entries
  - Page numbering
  - Table of contents

Credit: Chester County Historical Society

Administrative Metadata

- Manages resources:
  - File type
  - When/how it was created
  - Hardware / software
- Rights Management:
  - Copyright
  - Access and use restrictions
- Preservation:
  - Tracks preservation activities
  - Records preservation actions

Implementing Metadata

- Develop a style guide
- Determine workflow
  - Common Fields:
    - author/creator
    - copyright
    - credit
    - caption/description
    - keywords
    - title/object name
    - location related fields date created
    - date updated
    - original filename

Hardware, Software, Preparation

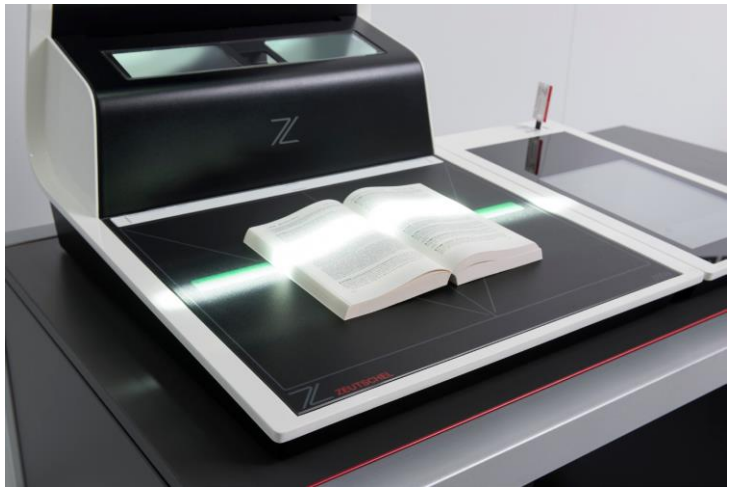
Choice of Equipment

- Suitable for materials
- High-res digital cameras for minute details
- Flatbed scanners for documents and photos
- Special book scanners
- Consider your budget

Types of Scanners

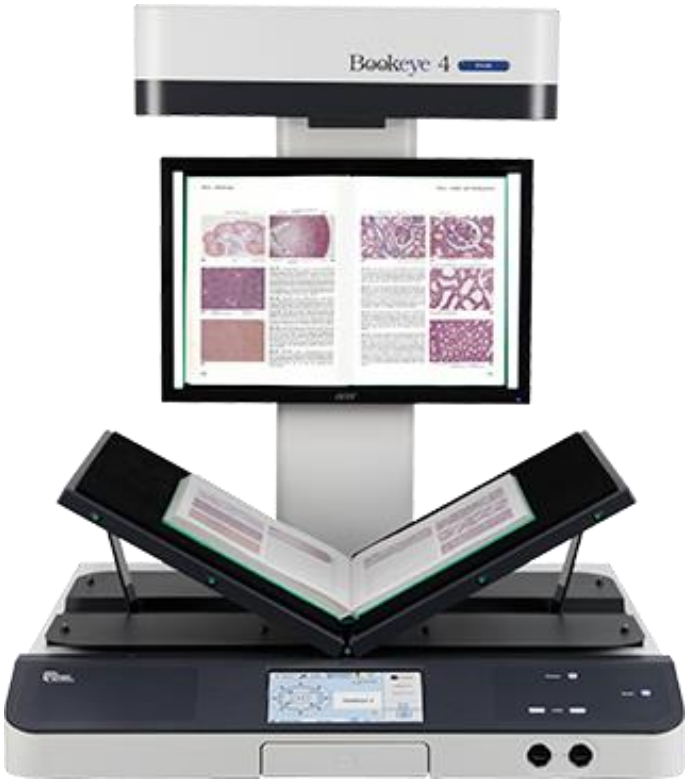
- Flatbed scanners
- Slide scanners
- Open book scanners
- Drum scanners
- Wide format scanners
- Digital cameras

Credit: http://largeformatscanners.com/book-scanners.htm



Various Epson and Zeutschel Scanners

Book Scanners



Credit: http://www.imageaccess.us/?page=WhitePapersBookScannerQualityCheck&lang=en / http://dtdch.com/dt-bc100/

Software

- Scanner software
  - Passes information to computer or editing software
  - Output in specific formats
- Image Editing software
  - Manipulates images, create derivatives
  - Requires training and upgrades
- Digital Asset Management software
  - Large number of files can be managed
  - Should have on-site

Digital Asset Management Systems (DAMS)

- Digital Asset Management refers to how you take in, handle and distribute everything you have in digital form, from digital images to word processing documents.

- DAM systems provide an infrastructure for preserving and managing digital assets.
  - Luna Insight
  - CONTENTdm
  - Islandora

Credit: Adapted from LYRASIS

Content Management Systems (CMS)

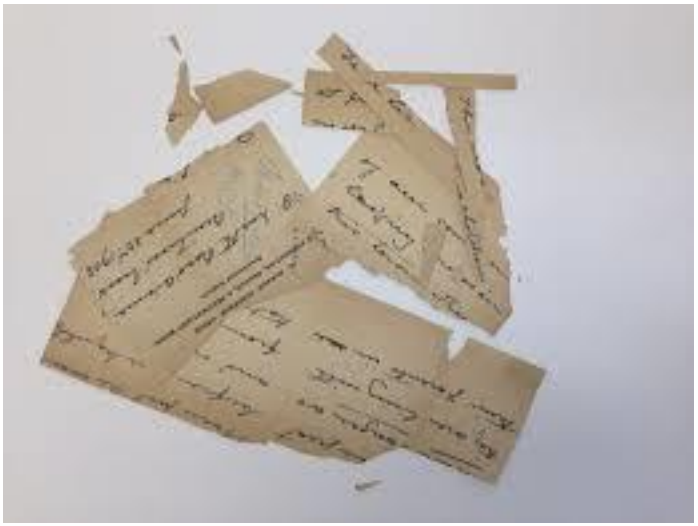
- A content management system (CMS) is a computer application used to create, edit, manage, and publish content in a consistently organized fashion.

- It manages information about the object and associated metadata.

Credit: Adapted from LYRASIS

Digitization Preparation

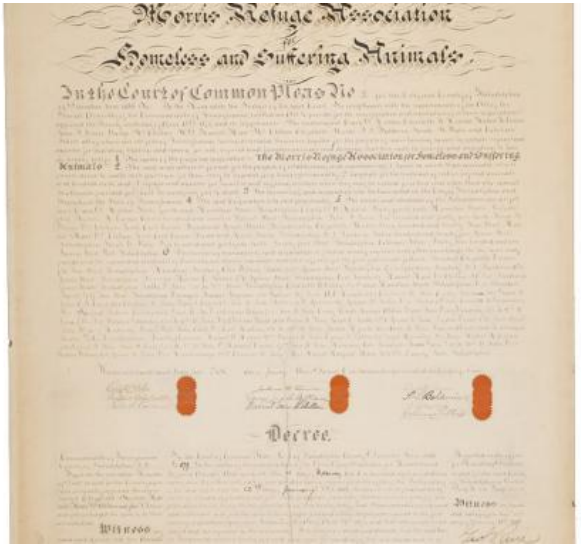
- Prepare documents
  - Assess condition
  - Remove hardware



- Support of Documents
  - Use both hands
  - Measure scanning bed
  - Do not leave unattended
  - Support books with cradles or wedges

Digitization Preparation

- Pages/Corners
  - Turn pages from fore edge
  - Avoid using moisture
  - Do not pinch corners
  - Unfold folded corners
  - Do not make new creases
- Seals
  - Very fragile
  - Do not add weight
  - Do not use glass



Digitization Preparation

- Keep Documents in Order
  - Contents should stay in sequence
  - One document at a time
  - Return documents to storage at end of day
- Annotation and Labeling
  - Do not labeling documents
  - Do not use sticky notes
  - Use paper strips

The Scanning Area

- Workstation should provide support
- Keep area clean
- No food or drink
- Pencils only
- No skin moisturizers
- Hands should be clean and dry

Gloves, Tools, and Cleaning

- Latex or Nitrile gloves only
- No handling aids
- No cleaning liquids
- Note any damage before scanning

Planning:
In-house vs. Outsourcing

In-house Approach

Pros:

- Learn as you go
- Retain control
- Provide security & proper handling
- Prioritize collection
- Maintain high quality

From the Cornell Digital Imaging Tutorial

In-house Approach

Cons:

- Large investment
- No set per-image cost
- Institution pays expenses
- Limited production facilities
- Range of staff expertise required
- No knowledge of best practices

From the Cornell Digital Imaging Tutorial

Outsourcing Approach

Pros:

- Cost containment, limited risk
- Costs typically lower
- Vendors handle large volumes
- Vendor pays all expenses
- Range of options and services

Outsourcing Approach

Cons:

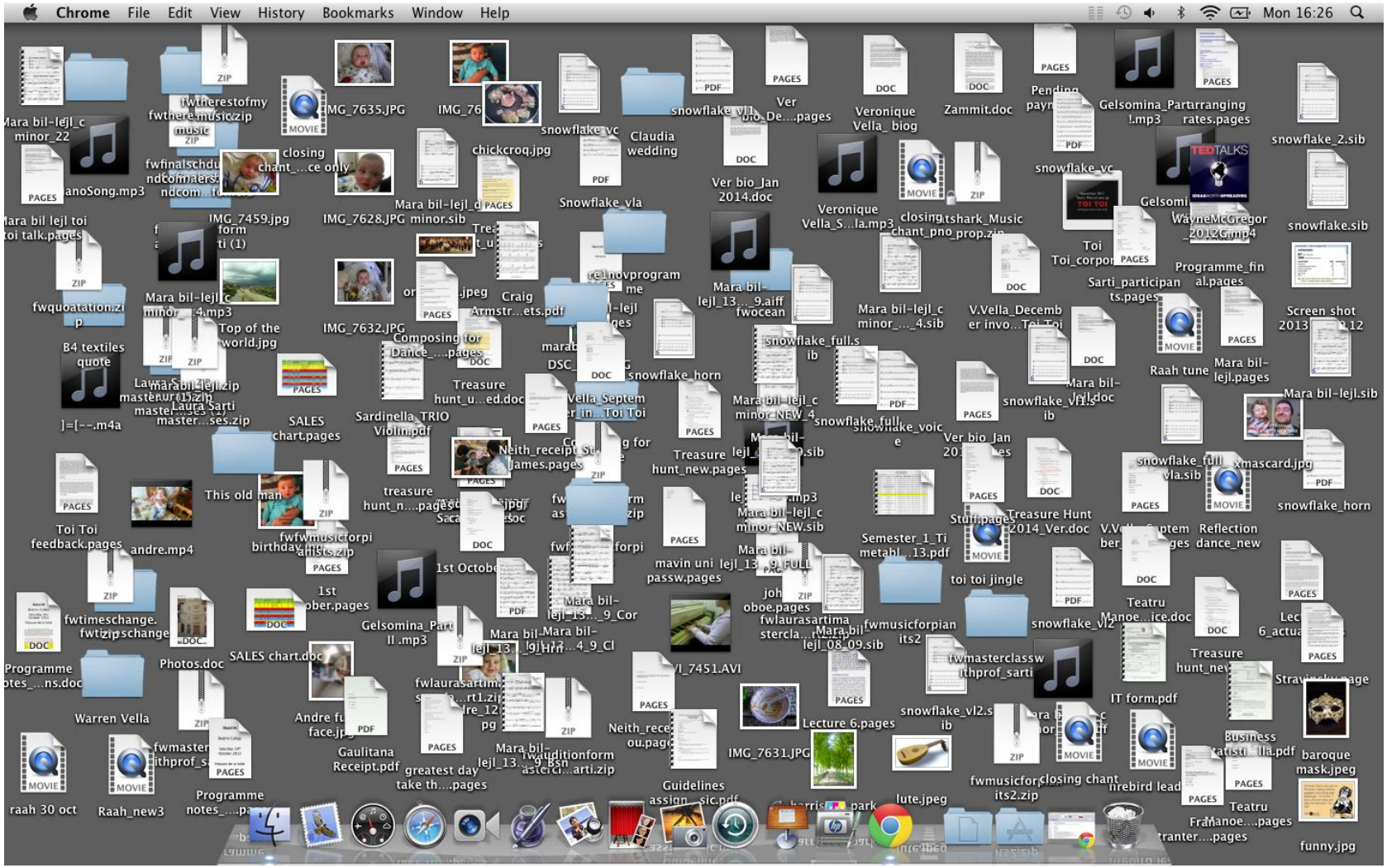
- Institution has less control
- Possible vendor instability
- Vendor inexperience cultural institutions
- Lack of best practices
- Challenges in communication
- Security and handling issues

From the Cornell Digital Imaging Tutorial

Long-term Management

Digital File Management & Storage

- Keeping track is not a simple task
- File management is high priority



Credit: http://www.galleryhip.com/

Importance of Digital File Management

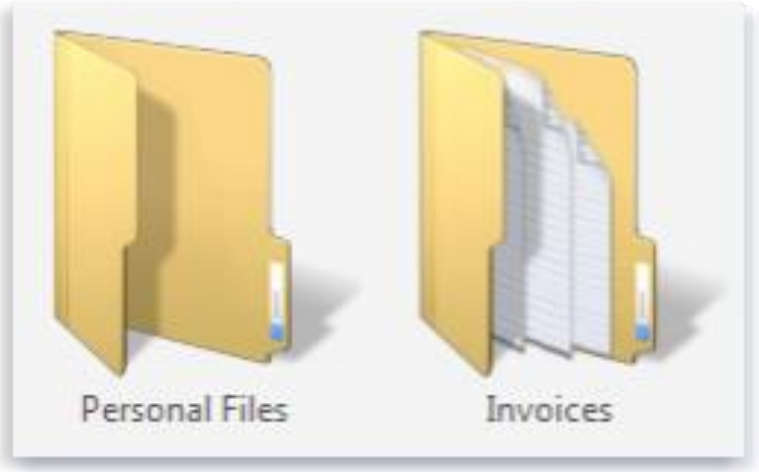
- Edited versions may get distributed
  - Track derivatives!
- Exist in multiple formats / media
- Lack of management = lack of protection

Benefits of Digital File Management

- Retrieve easily
- Restrict access
- Short-term and long-term storage
- Keep a safe copy
- Original version safely stored

Organizing Digital Files

- Be descriptive
  - Easy to remember
  - Not too complicated
- Be consistent
  - Develop a technique
- File management software



File Naming

- Standardize the names
- Use standard file extensions
- Short and simple
- Start with the year or date
- Include descriptive text
- Avoid using spaces or special characters

- Example: 2016_06_CCAHA_MarketingPhotos_01

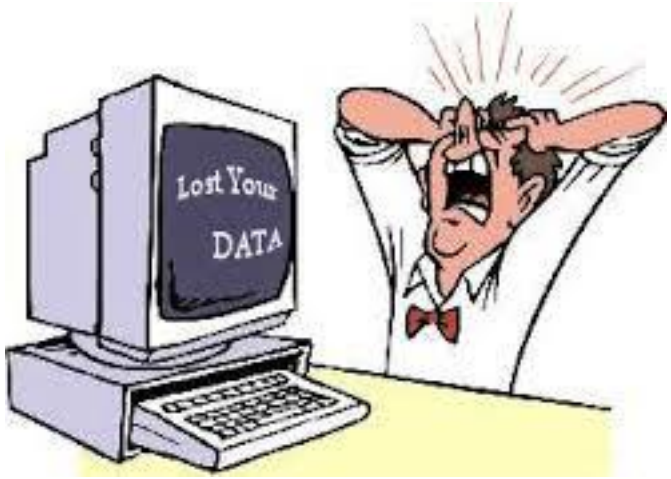
File Naming Resources

- Stanford University Libraries, File Naming Best Practices https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming

- University of Illinois, Best Practices for File Naminghttp://guides.library.illinois.edu/c.php?g=348391&p=2347452

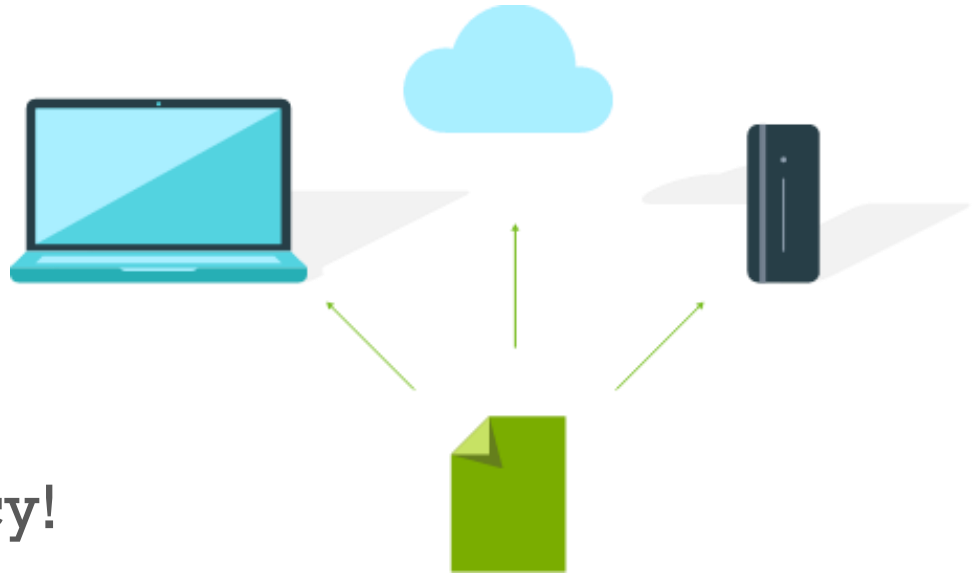
Backing Up Files

- Without a proper backup, you can lose everything
  - Computer failure
  - Mistake in editing
  - Natural disasters



Best Practice for Backing Up

- The 3-2-1 Rule
  - At least **3** copies
  - On **2** different formats
  - With **1** copy off-site



It's all about Redundancy!

Best Practice for Backing Up

- Original on computer/server
- Subsequent Backups
  - External hard drives
  - Online/Cloud storage
- Separate Locations

Credit: http://www.uwphotographyguide.com/protect-digital-images-external-hard-drive

Digitization Summary

- Mission
- Focus on goals
- Audience
- Prioritize
- Use best practices
- Manage your assets
- Backup, backup, backup!

Questions…

DHPS | NY

Documentary Heritage & Preservation Services

FOR NEW YORK

Tamara Talansky
Preservation and Imaging Project Coordinator
Conservation Center for Art & Historic Artifacts

Email: ttalansky@ccaha.org

DHPS | NY

Documentary Heritage & Preservation Services

FOR NEW YORK

Anastasia Matijkiw
Program Coordinator

264 S. 23rd Street
Philadelphia, PA 19103
t 215-545-0613 *f* 215-735-9313
e amatijkiw@dhpsny.org

Jason Henn
Communications Assistant

264 S. 23rd Street
Philadelphia, PA 19103
t 215-545-0613 *f* 215-735-9313
e jhenn@dhpsny.org