

# Digital Preservation

*Starting and managing a digital preservation program*

Mary Wallace, MA, Audiovisual Archivist, Walter P. Reuther Library, Archives of Labor and Urban Affairs

Arjun Sabharwal, MA, MLIS, Digital Initiatives Librarian, William S. Carlson Library, University of Toledo  
October 18, 2013

# Why digitize?

- To Save space in your library
- To preserve information on older formats
- To make that information more accessible to more people
- To make your institution more discoverable

# Assessing your needs

- Audience (affects the scope of collections, metadata schema, subject classification)
- Collections
- Technology and equipment
- Administrative support (budget, staffing)

# Equipment

# Equipment needs depend on what's being digitized

- Images and Document need image-capture devices like scanners and digital cameras
- Analog Audio and video need media players
- Film needs certain types of projectors and digital video capture devices

# Turning analog photos, books, or papers to digital

- Book scanner
- Form-feed scanner/copier
- Regular photo scanner



# Turning Analog to Digital

- Digitizing of fragile papers and photographs must be done on a flatbed scanner—not form-fed.



- Larger Objects can be digitized with a digital camera and special tripod

# Turning analog audio to digital

- Reel-to-reel player
- Record player
- Cassette player
- CD players
- Cables





# Analog Audio to Digital

- Make sure equipment has an audio output somewhere.
  - RCA Outputs
  - Headphone output (Adapter?)



# Turning analog video to digital

- VCR
- U-Matic player
- Professional Betacam player
- Other types or players?
- Analog to digital converter
- VHS to DVD combo



# Analog film to digital

- Telecine machines run the film through a projector with an integrated video recorder to convert it into a video format



- Film transfer screens allow you to project the film on a screen which enables the image to be captured by a digital camcorder



Software

# Production Software Needed

- Adobe Acrobat for scanning and creating PDFs
- Photo manipulation software - PhotoShop
- Digital Photo asset management software - Adobe Bridge
- Scanning Software - Silverfast
- Format conversion software (open source):  
FormatFactory (videos also); Spesoft Image Converter (still images); Audacity (audio); Sound Forge (audio)

# Storage

# Storage

- Optical Media – CDs and DVDs
  - CDs-700 MB; DVDs-4.7 GB
  - Keep in stable environment (20-50% RH; 40-73° F)
  - Do not mark on them
  - Handle by the outer edges or the center hole
  - Keep out of direct sunlight and UV light
  - Store vertically in separate cases or envelopes

# Storage

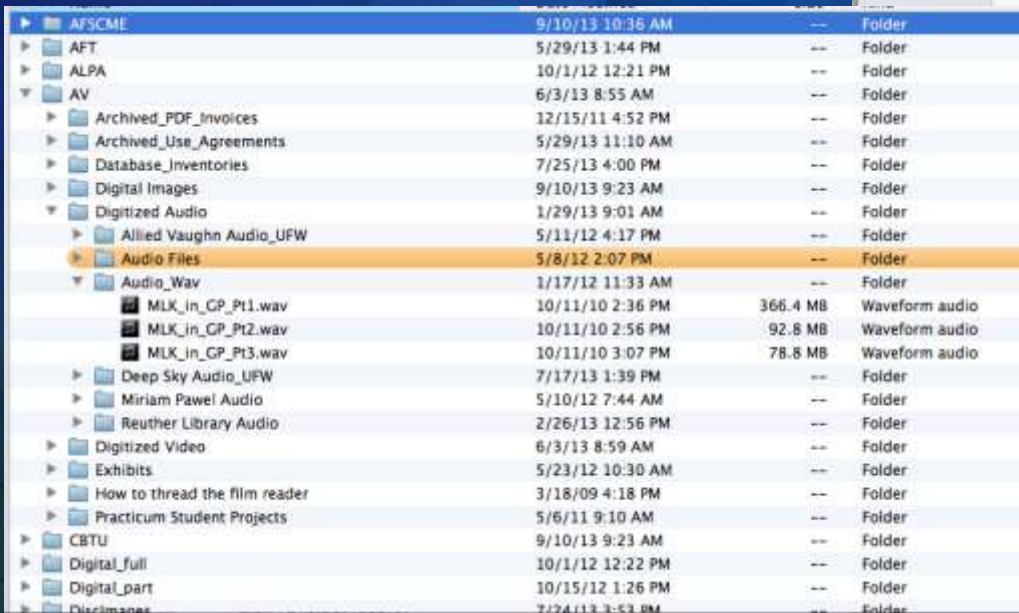
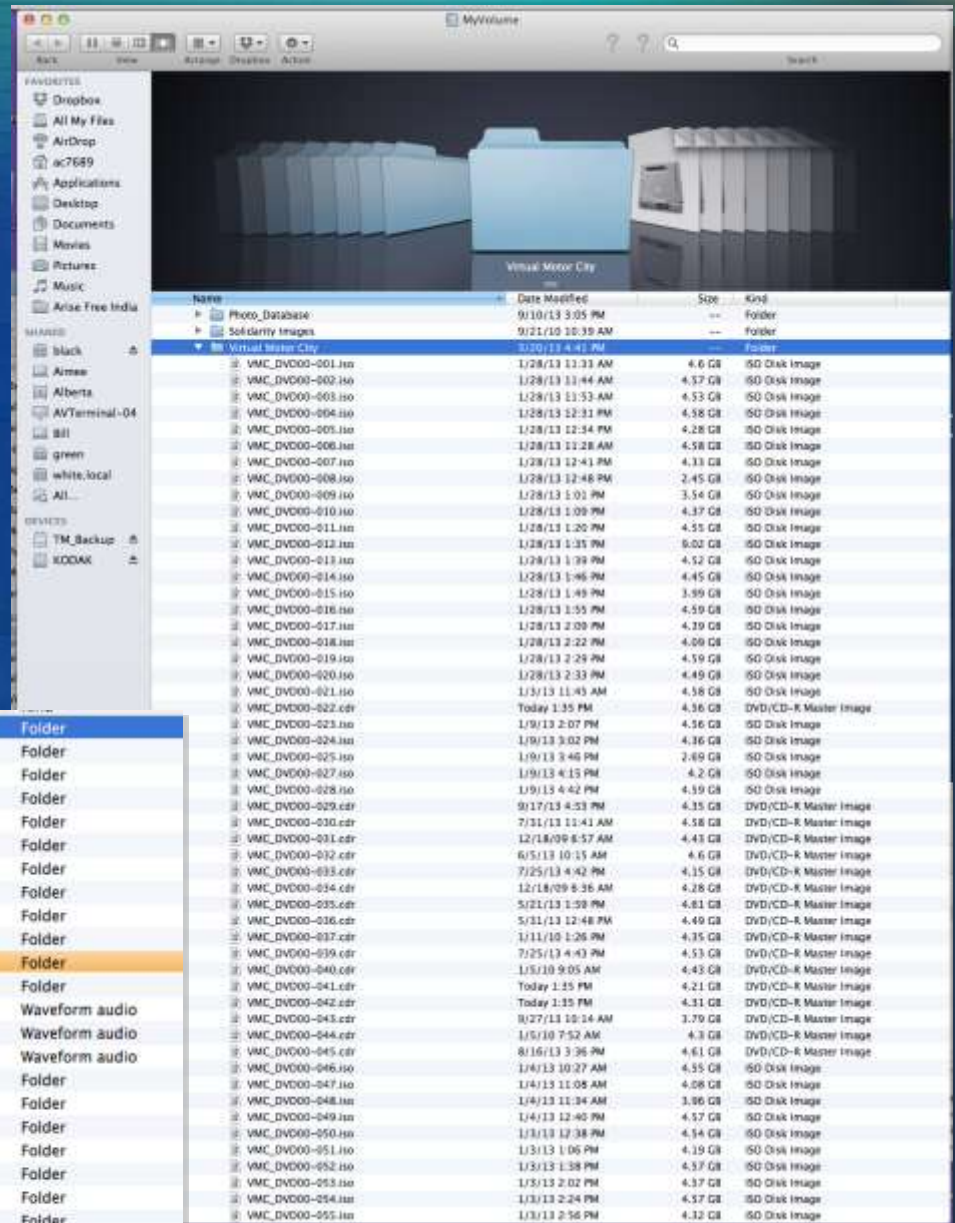
- Servers

- Multiple drives to run multiple back-ups
- Expandable RAM
- Uninterruptable power supply (UPS)
- Climate controlled room (65-80° F)



# Storage

- Servers with disc images of each optical disc
- Server with preservation copies of digitized files



# Digital Repositories

- Digital Libraries:
  - Open-source: DSpace, Omeka, KORA, Fedora, Greenstone
  - Commercial: CONTENTdm, Luna
  - Pros and cons
- Best practices: digital reformatting, procedures, specifications, metadata, workflow, digital curation
- Backup, disaster planning

# Best Practices

# Formats and Settings

- Scan documents to PDF. Run OCR text recognition and convert to PDF/A
- Scan as a TIFF file at 600 ppi at 100% size, 48→24 bit Color
- Image on Access Interface (website) should be 72 dpi
- Preservation audio files should be WAV files and access versions of Audio files can be MP3
- Preservation video files can be AVI or QuickTime format and the compressed files can be MPEG.  
50 seconds of uncompressed video = 1 GB storage.

# Scanning

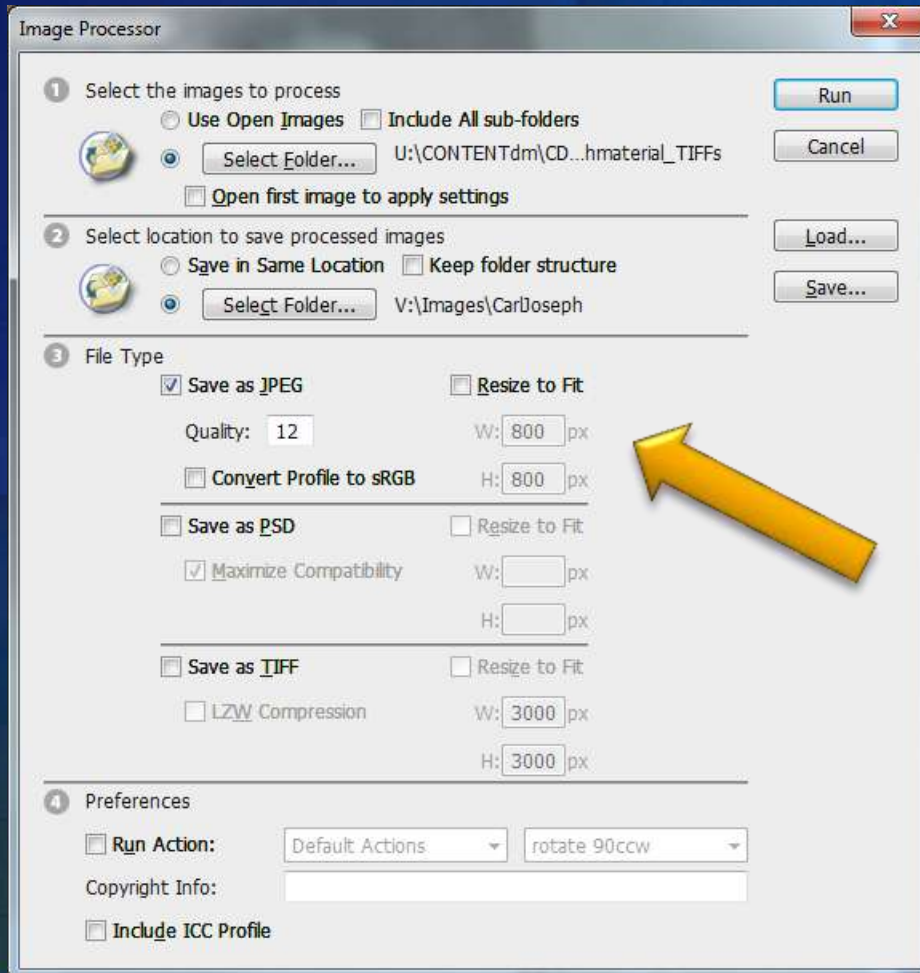
# Scanning in Photoshop

Size (px, in, cm/mm)/resolution (ppi)



# Batch processing in Photoshop

Time-saving solutions for mass-digitization projects



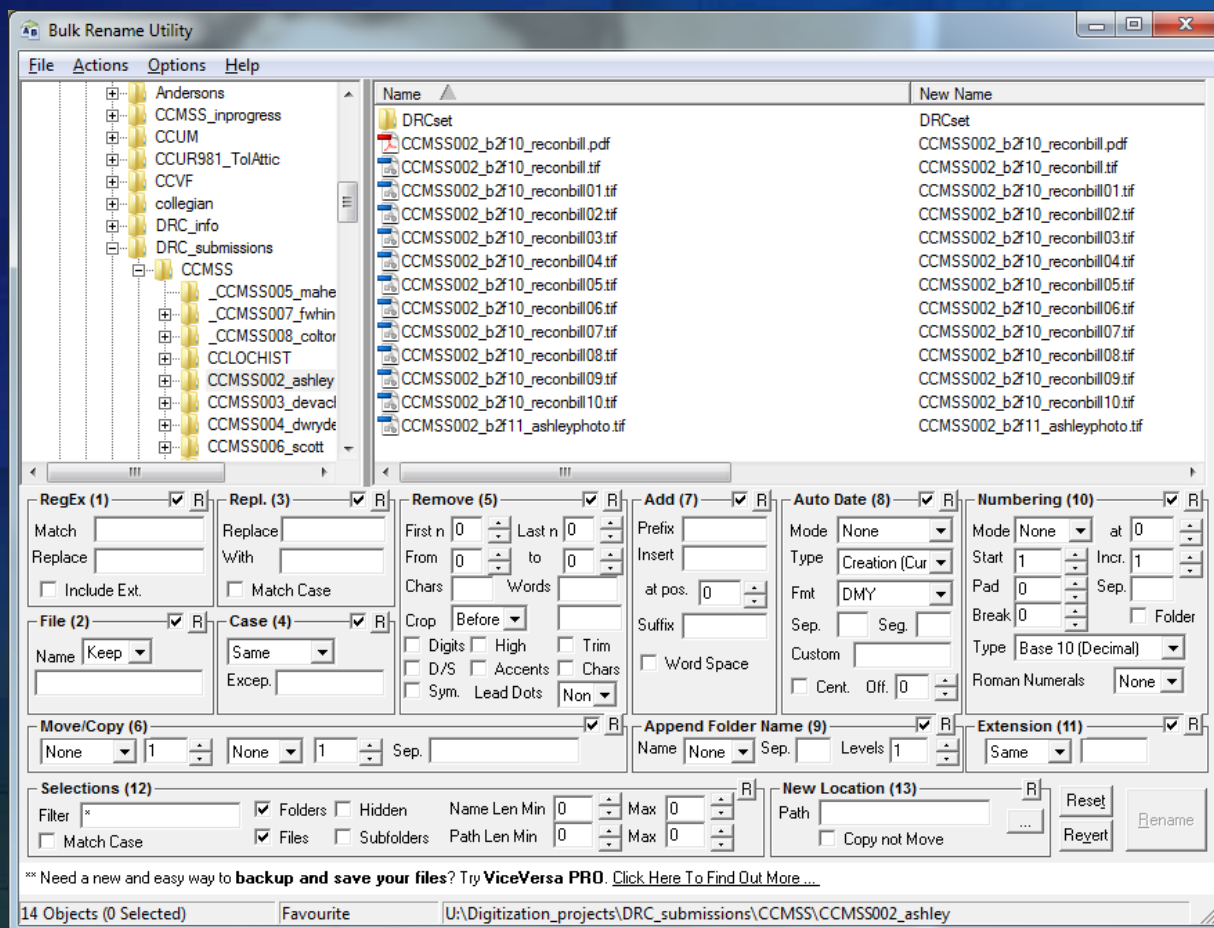
Conversion (from  
TIFF to JPG

Run various  
actions (macros)  
Batch resize

# Bulk rename utility (free software)

Time-saving solutions for mass-digitization projects

Bulk renaming  
Seriating  
Auto-dating





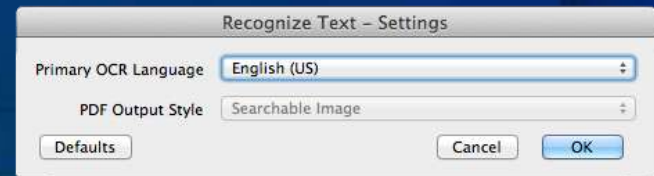
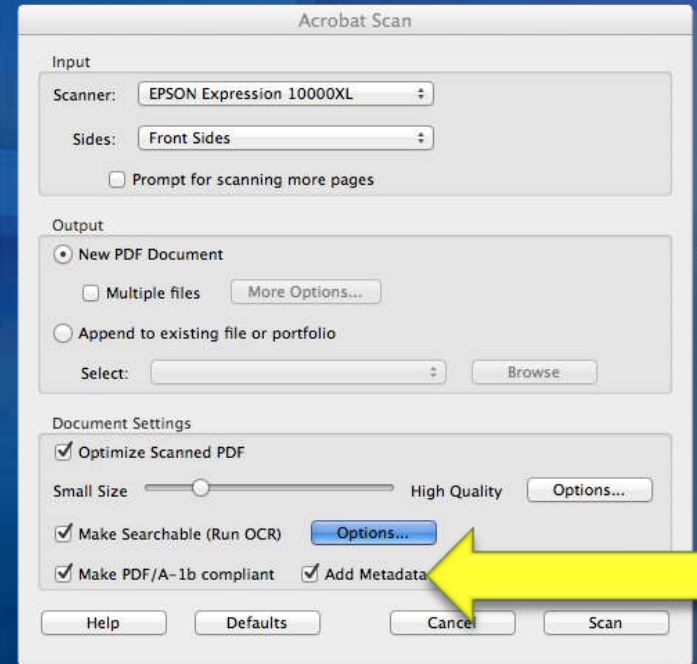
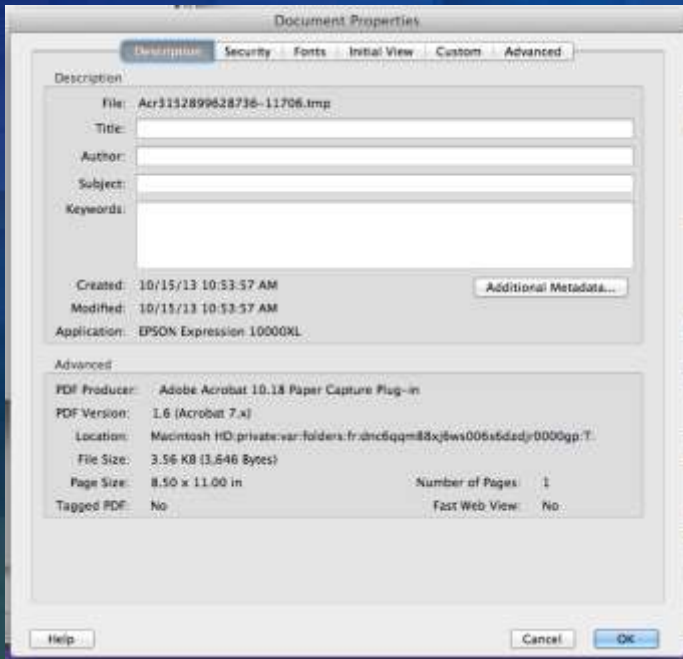
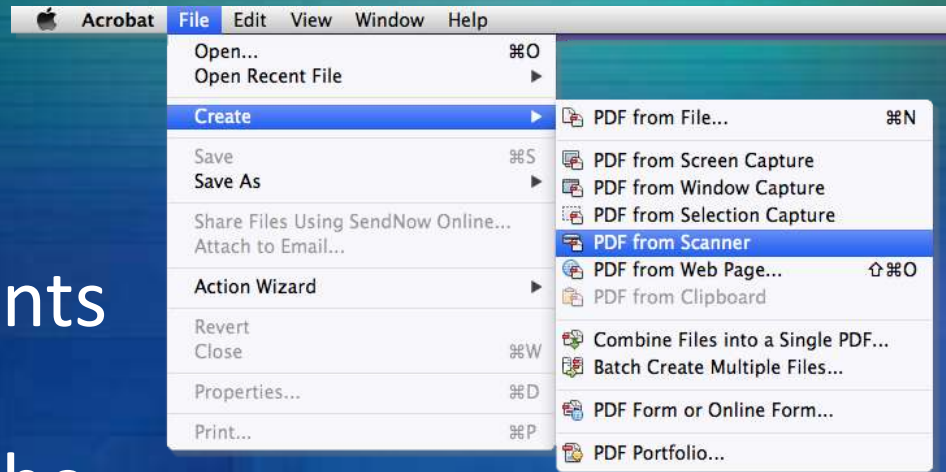
# Metadata, Discovery, Information Retrieval

- Choosing the appropriate schema
  - Dublin Core, VRA, JurMeta, CDWA (Getty)
  - MARC
- Classification systems:
  - Library of Congress, Dewey, MeSH
  - Hicks, Schiller (law materials)
- Best practices: digital reformatting, procedures, specifications, metadata, workflow, digital curation, migration, backward compatibility
- Promoting your collections

# Metadata

# Metadata

- Metadata can be added to documents through Adobe Acrobat while in the scanning process



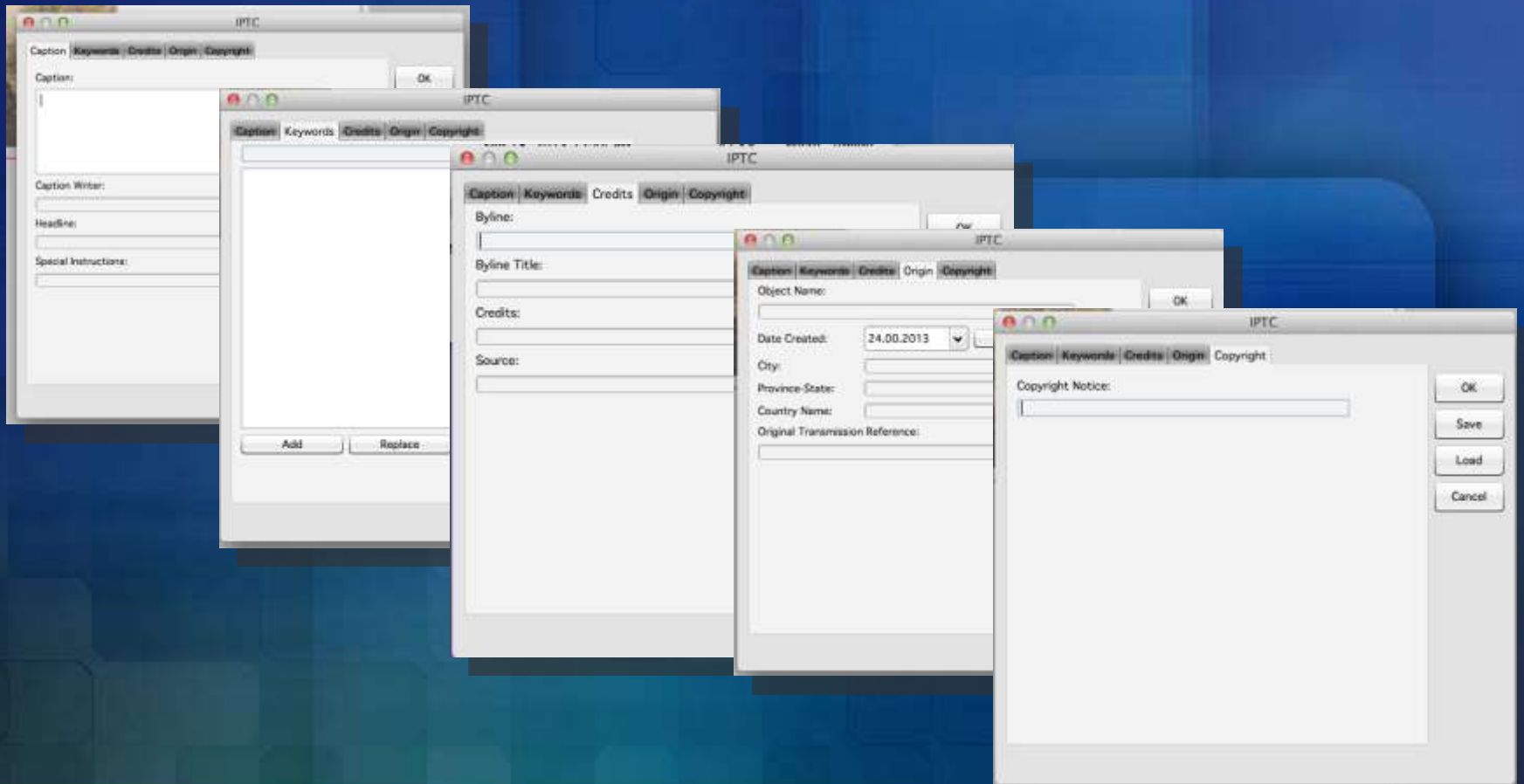
# Metadata

- Metadata can be added to images while in the scanning process.



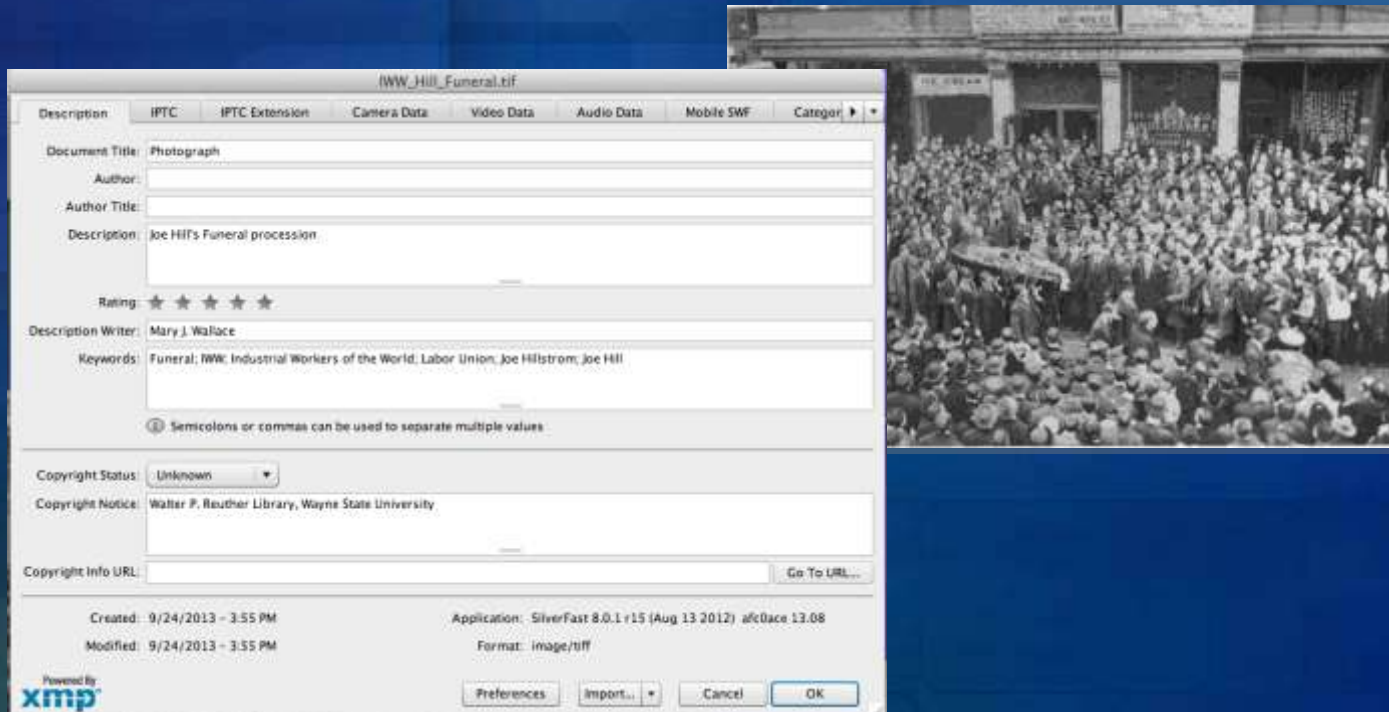
# IPTC Fields map to Dublin Core

- Caption, Keywords, Photographer or creator of Document, Object Name, Copyright Notice



# Metadata can be added to each specific digital object

- PhotoShop allows metadata to be added via the File → File Info window



# Metadata can be added as a batch

- Adobe Bridge allows metadata to be added to a group of digital objects, by allowing the user to “create” a metadata template

Create Metadata Template

Template Name:

Choose the metadata to include in this template:

**▼ IPTC Core**

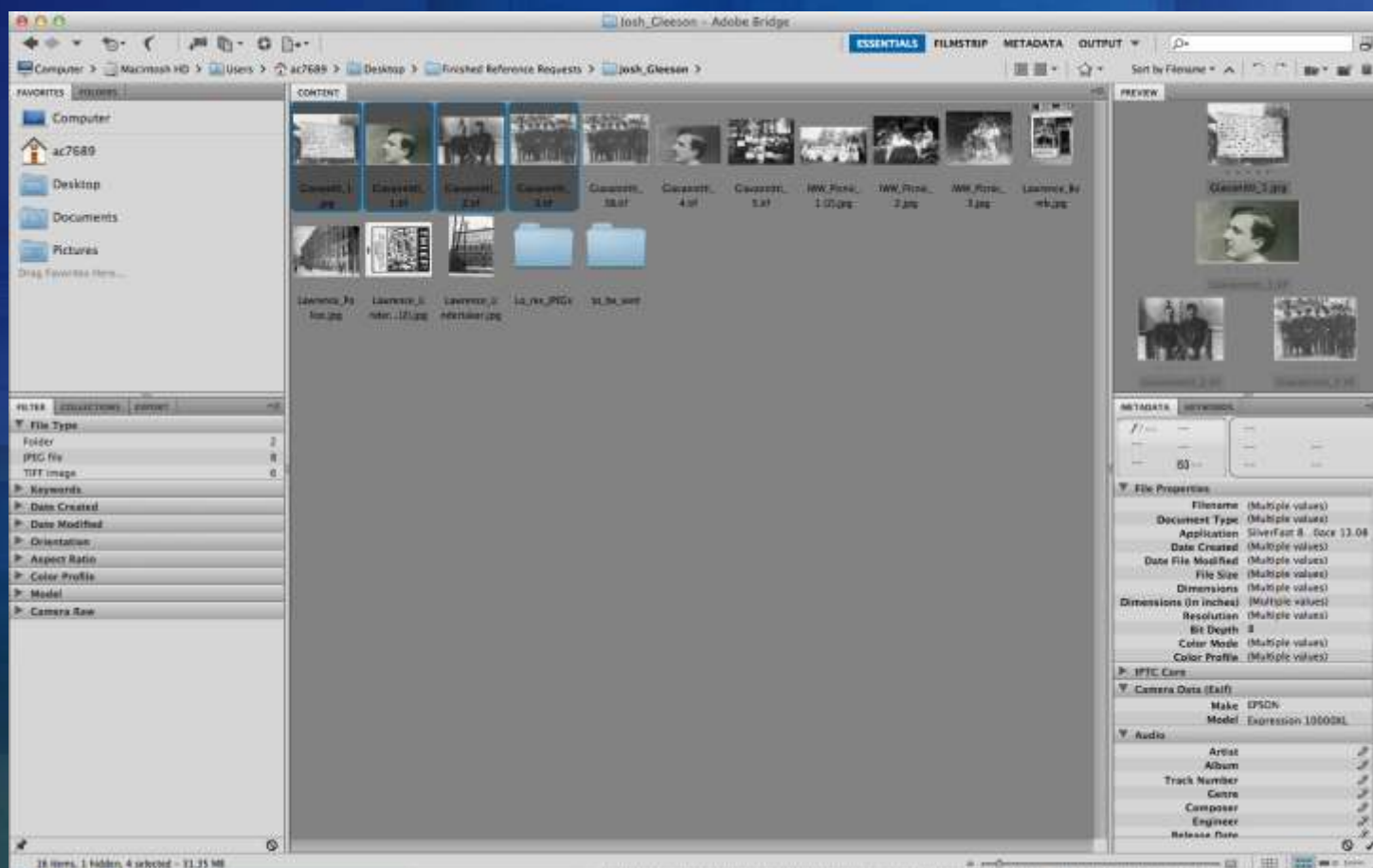
<input checked="" type="checkbox"/>	Creator	: Industrial Workers of the World (IWW)
<input type="checkbox"/>	Creator: Job Title	:
<input type="checkbox"/>	Creator: Address	:
<input type="checkbox"/>	Creator: City	:
<input type="checkbox"/>	Creator: State/Province	:
<input type="checkbox"/>	Creator: Postal Code	:
<input type="checkbox"/>	Creator: Country	:
<input type="checkbox"/>	Creator: Phone(s)	:
<input type="checkbox"/>	Creator: Email(s)	:
<input type="checkbox"/>	Creator: Website(s)	:
<input type="checkbox"/>	Headline	:
<input checked="" type="checkbox"/>	Description	: Arturo Giavonnitti
<input checked="" type="checkbox"/>	Keywords	: Industrial Workers of the World, IWW, Giavonnitti, Italian, Joe Ettor, Joseph Caruso, Sermon on the Common, trial
<input type="checkbox"/>	IPTC Subject Code	:
<input type="checkbox"/>	Description Writer	:
<input type="checkbox"/>	Date Created	:
<input type="checkbox"/>	Intellectual Genre	:
<input type="checkbox"/>	IPTC Scene Code	:
<input type="checkbox"/>	Sublocation	:
<input type="checkbox"/>	City	:
<input type="checkbox"/>	State/Province	:
<input type="checkbox"/>	Country	:
<input type="checkbox"/>	ISO Country Code	:
<input checked="" type="checkbox"/>	Title	: Industrial Workers of the World (IWW): Arturo Giavonnitti
<input type="checkbox"/>	Job Identifier	:
<input type="checkbox"/>	Instructions	:
<input checked="" type="checkbox"/>	Credit Line	: Walter P. Reuther Library, Archives of Labor and Urban Affairs, Wayne State University
<input type="checkbox"/>	Source	:
<input type="checkbox"/>	Copyright Notice	:
<input checked="" type="checkbox"/>	Copyright Status	: Public Domain
<input type="checkbox"/>	Rights Usage Terms	:

! Only checked properties will be added/changed to this template.

Properties selected: 6

# Batch Metadata

- That Metadata Template can be applied to a group of digital objects





# Batch Metadata

- Once the Template is applied to all selected digital objects, they all contain the same metadata. Individual fields can be edited later to add data that is specific to that object.
- Templates can be used over and over and can be modified by “editing” the template in Adobe Bridge



# Batch Metadata

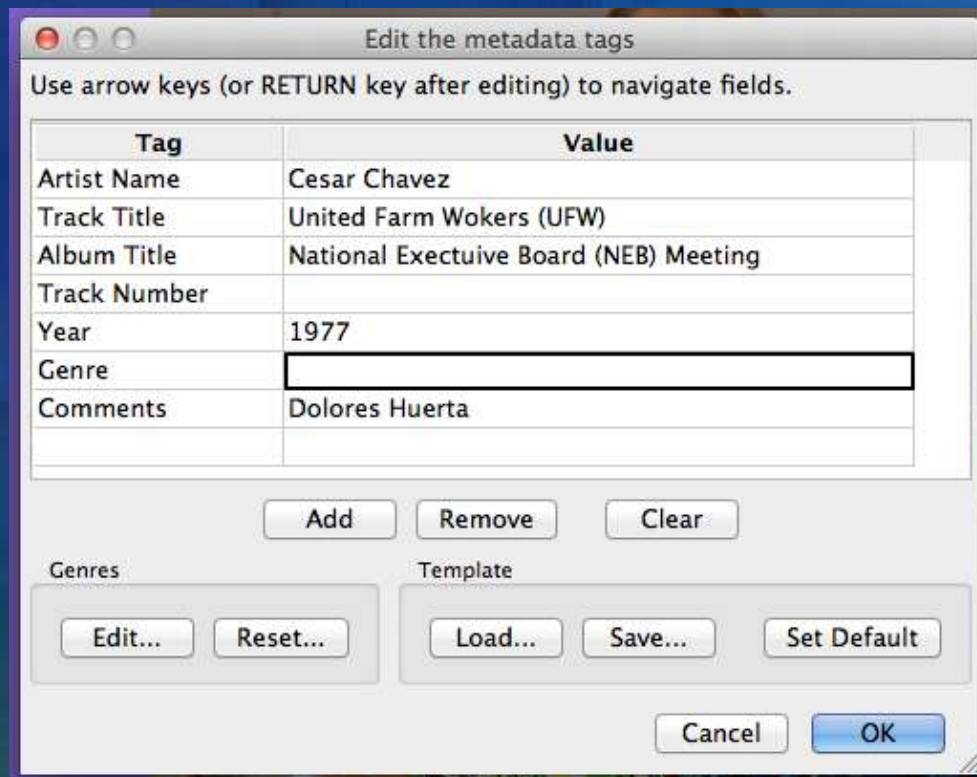
## Excel worksheets

The image displays three overlapping windows illustrating a batch metadata workflow:

- Microsoft Excel:** A spreadsheet titled 'botminutes1872-1918.xls' with columns for Title, Description, Date Created, Document Type, Physical Dimer, File Type, and Collection. It lists various historical documents from the University of Toledo.
- University of Toledo Libraries:** A search result for 'Section on Surgery' with details on the author (The Academy of Medicine of Toledo and Lucas County) and a list of files including 'CONGRESS\_BIBLIOGRAPHY.pdf'.
- Image Description:** A metadata record for an image titled 'We're all going downtown', showing details such as description, date created (1933-03-02), file type (image/jpeg), and subject (Business and Commerce, Municipal Government).

# Metadata in other formats

- Audio Metadata can be added in Audacity



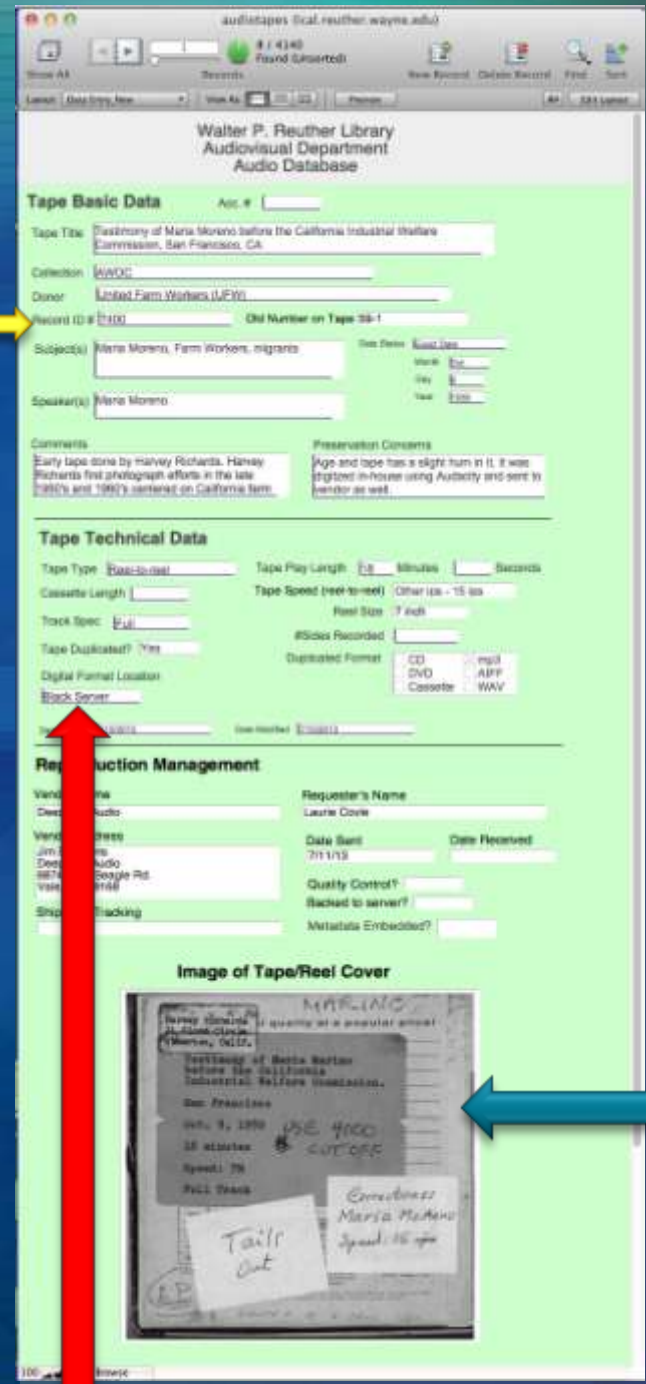
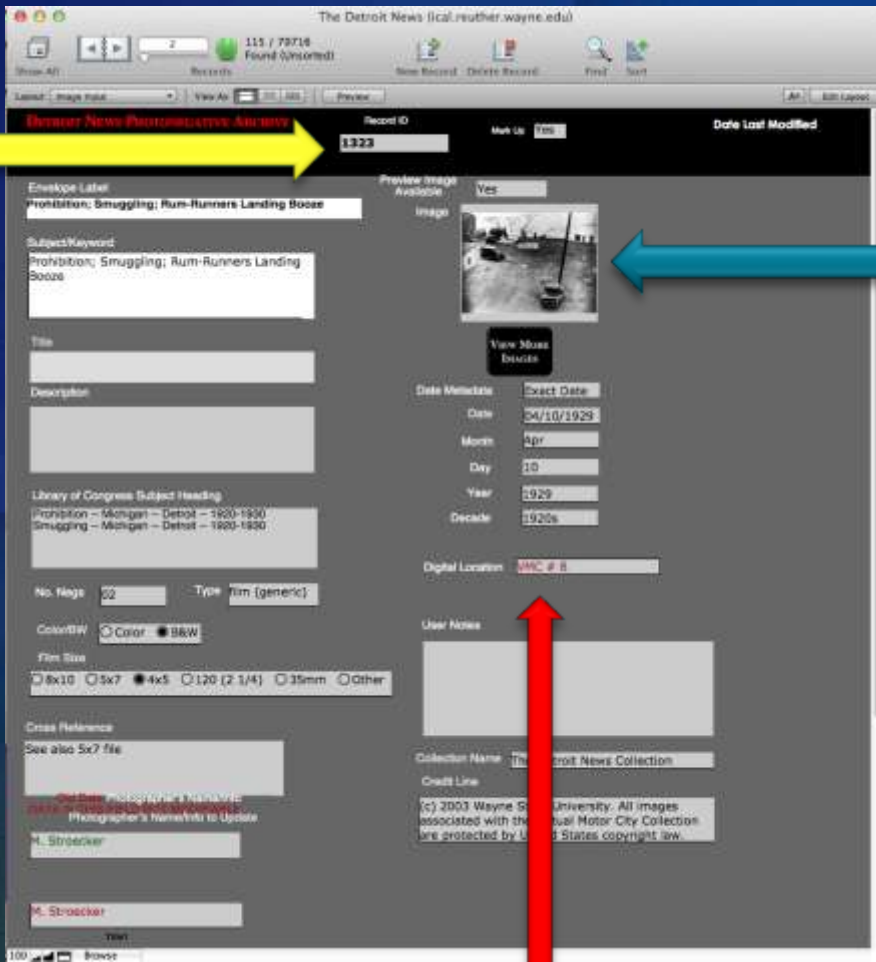
# Internal Data Controls

# Internal Access Systems

- Database software like FileMaker Pro or Microsoft Access
- Record ID number matches the scanned item's name
- Container field shows photo of object
- Record shows where digital version can be found
- Database can be published on the Web, or exported to a spreadsheet program like Excel

# Internal Access

- Image and Audio database examples



# Promoting Your New Digital Assets

# Promoting your new digital assets

- Blog on your institution's website

- With a link to the digital collection online



The screenshot shows the Reuther Library website with a prominent announcement for the "Antecedents to Education Reform Historical Collection". The announcement includes a photograph of a protest with signs that read "KIDS ARE The REAL LOSERS" and "Accountability with a Class of 4000". The text describes the collection as a digital archive of nearly 30 years of historical records related to the American Federation of Teachers (AFT) and education reform.



This screenshot displays the American Federation of Teachers website, specifically the page for the "Antecedents to Education Reform Historical Collection". It features a search bar, navigation links, and a featured document titled "Letter to Jimmie Jackson from Marilyn Raith".



The screenshot shows a digital document titled "Letter to Jimmie Jackson from Marilyn Raith" from the American Federation of Teachers (190-04). The document is dated 1960 and is a typed letter. The header includes the AFT logo and the title of the collection. The body of the letter discusses the AFT's stance on education reform and the role of teachers.



# Promoting your new digital assets

- Publications like newsletters – print and electronic



## Partnering with the WSU Library System

*To Make Resources Available on the Web*

The American Federation of Teachers (AFT), one of the Reuther Library's major donor unions, was frustrated by the narrative in the media that teachers' unions existed solely as obstinate impediments to education reform. AFT's history tells a very different story, as for years its members and leaders have initiated reforms in the areas of peer review, career development, wages, and shared decision-making and management. With its archivist, Dan Golodner, AFT decided to create a Web site to host historic, primary

documents from its vast archives at the Reuther Library that illustrate AFT-led efforts at education reform. This site would provide documentary evidence in the debate over education reform.

In March 2010, Golodner reached out to the Wayne State University Library System's Digital Library Initiatives (DLI) team for help. He described the project's goals to the DLI's Jon McGlone.

Together, they felt this was a good project to test two things: collaboration between the Reuther and the Library System; and Omeka, a new open source platform for displaying Web exhibits and digital collections in which the Library System was interested.

Golodner selected documents dating back to the early 1980s from the national headquarters' records and from the collections of AFT affiliates, as well as from various AFT-related personal collections. He then hired Archives Technician Jen Dawson to scan the documents. Dawson worked with the DLI to add metadata to the documents, and McGlone implemented Omeka for displaying them. The result can be viewed at <http://dlx.lh.wayne.edu/aft/>.

The collaboration with the Library System proved successful, and the Reuther Library formed an ad hoc Digital Projects Committee to standardize and streamline procedures for working with the DLI to create more such projects utilizing the Reuther's collections. The Committee decided that another test case was required to hammer out these policies,

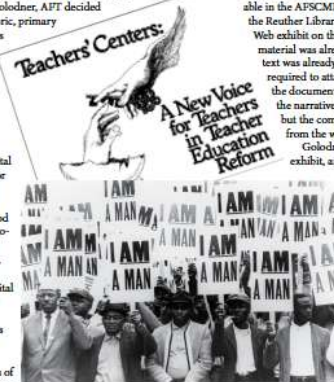
and chose to create a Web exhibit about the 1968 Memphis Sanitation Workers Strike led by the American Federation of State, County and Municipal Employees (AFSCME), another of the Reuther Library's major donor unions. It was at this strike that Martin Luther King, Jr. was assassinated.

This topic was chosen not only because of its general interest and the supporting documentation available in the AFSCME collection, but also because the Reuther Library had created a physical and Web exhibit on this topic in 2003. Much of the material was already digitized, and much of the text was already written. Significant work was required to attach appropriate metadata to the documents, add information to update the narrative, and address copyright issues, but the committee had a nice head start from the work done in 2003.

Golodner, who had curated the 2003 exhibit, and Johanna Russ, AFSCME

Archivist, worked together to select images and documents, describe them with metadata, and enhance the exhibit with a resources list, an introductory essay, and video clips from a 2003 symposium hosted by the Reuther. Golodner and Russ then worked with the Library System's Meghan Finch and Joshua Neels-Fox on metadata issues and site design. The Web exhibit *I AM A Man* went live in April 2012 on <http://dlx.lh.wayne.edu/iamaman/>.

Significant restructuring in the Library System and a director search at the Reuther has delayed future projects through the summer, but both sides are committed to working together in the future. The committee has many ideas for further Web exhibits and digital collections and is excited about the possibilities this new partnership has opened. ■



# Promoting your new digital assets

## ● Virtual Exhibitions



### Medicine on the Maumee: A History of Health Care in Northwest Ohio Virtual exhibit

In March 2012, the *Ward M. Canaday Center* opened its most recent exhibition on the history of medical practice, medical education, nursing, and hospitals in Northwest Ohio. The virtual exhibit is a digital reproduction of that exhibition ending in December 2012. It features materials from the Canaday Center's manuscript collections as well as others on loan from institutions that loaned them. Permissions were obtained to photograph and display them here. In addition to the materials in the exhibition, visitors can view videos, use the interactive timeline, and tour the exhibition area virtually. This site also provides links to research resources at various special collections, archives, and libraries to inspire and support individual research. Please address questions specific to the virtual exhibition to [Aqun Sahbanyal](#), the Digital Curator of the virtual exhibit.

#### Quick links:

- [View the exhibition now](#)
- [Navigating the virtual exhibition](#)
- [Historical Timeline](#)

© 2011 University of Toledo, Ward M. Canaday Center for Special Collections. All rights reserved.

[Contact Us](#) - [Site Map](#)



# Promoting your new digital assets



Facebook profile page for the Walter P. Reuther Library of Labor & Urban Affairs. The profile picture is a black and white portrait of a man in a suit. The cover photo is a large, modern glass-walled building. The page shows 1,010 likes and 91 people who were here. Navigation tabs include Library, Photos, Likes, Map, and Notes. A search bar is visible at the top.

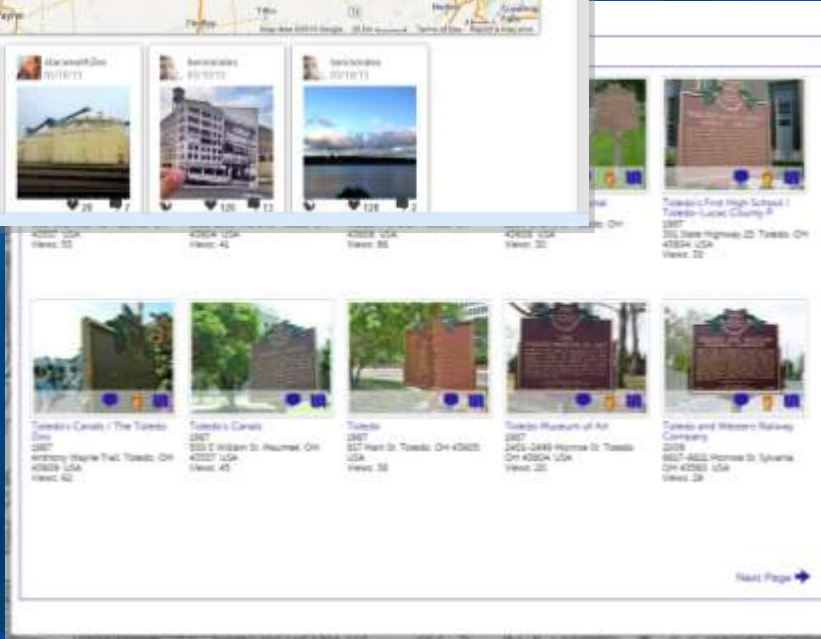
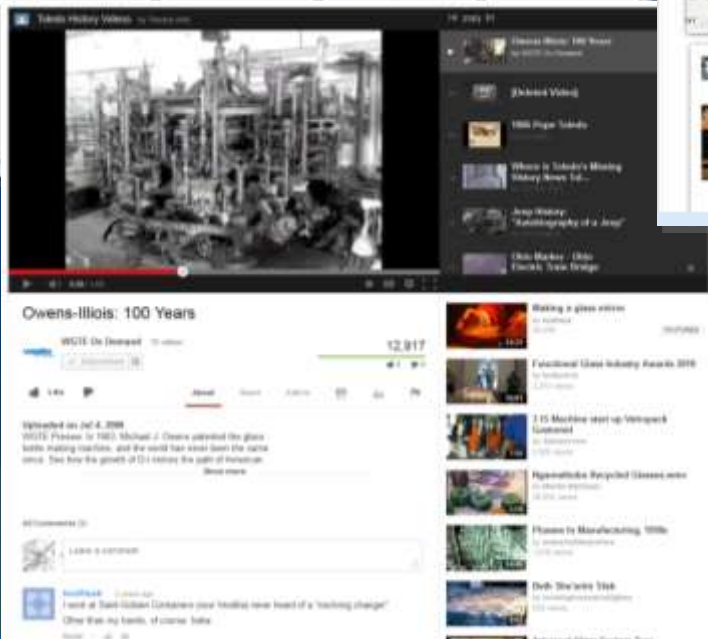
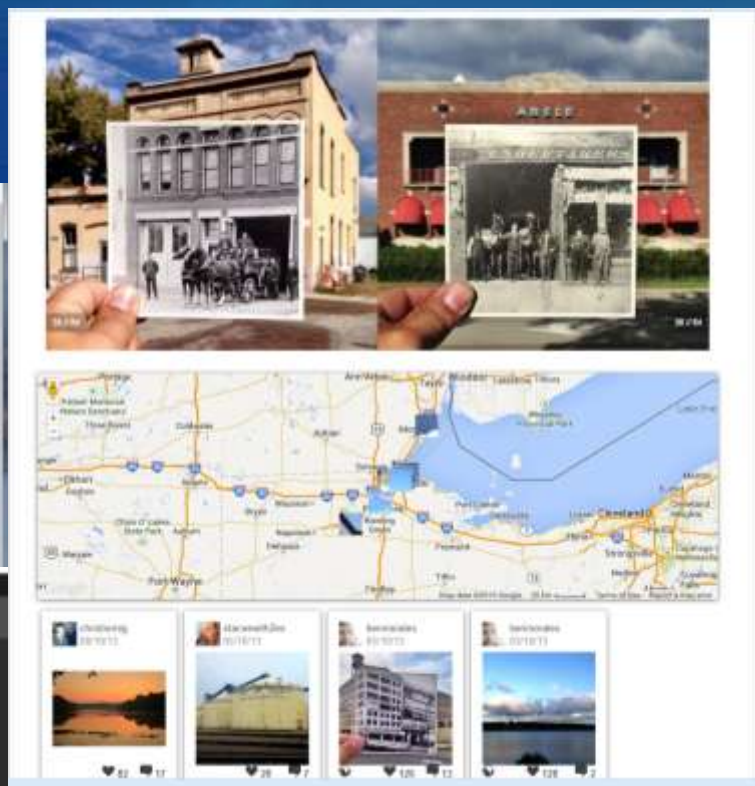
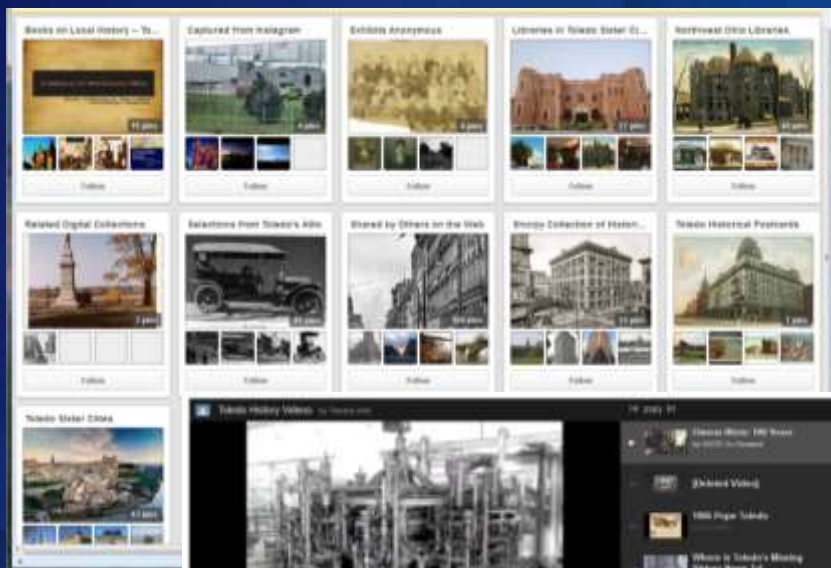


Facebook post from the Walter P. Reuther Library of Labor & Urban Affairs. The post is a share of a link, dated August 23. The text of the post reads: "Fun Photo Friday, March on Washington edition: (For more on the historic 1963 march, see: <http://reuther.wayne.edu/taxonomy/term/1646>)". Below the text is a black and white photograph of a large crowd at a demonstration, with a man in the foreground raising his fist. The caption for the photo reads: "Walter P. Reuther Library (25406) Civil Rights, Demonstrations, 'March on Washington,' King, 1963 reuther.wayne.edu". The post has 6 likes and 2 comments.

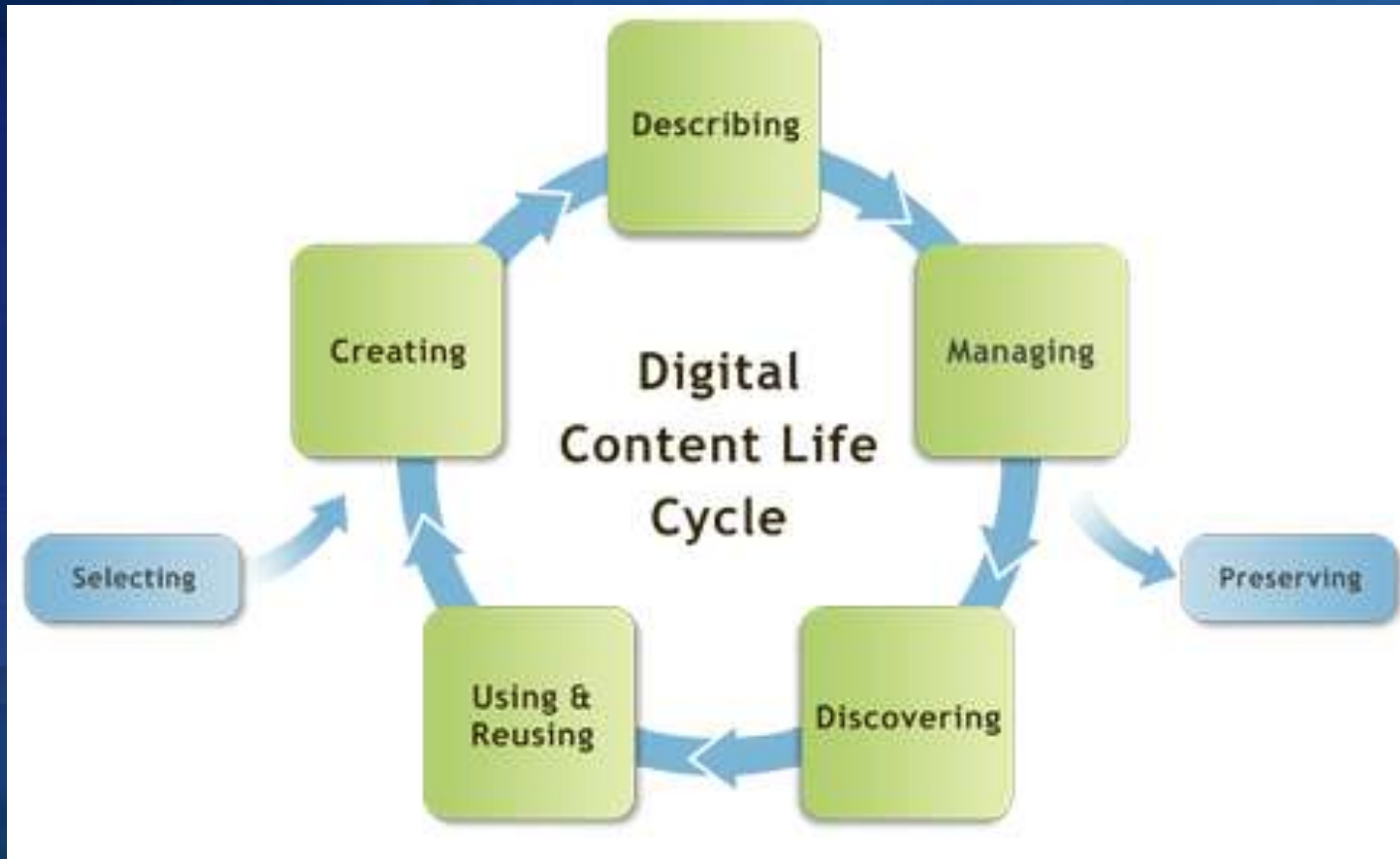
## ● Social Media posts

# Promoting your new digital assets

## ● Social Media



# Digital content lifecycle



Source: Life Cycle Models for Digital Stewardship, Library of Congress

...just a word about Copyright

# Copyright in the Archives

- Deed of Gift addresses copyright

*“In addition, the parties agree that:*

- 1. The Donor transfers ownership of his/her records or papers to the Repository, including all copyright and literary rights, which the Donor holds in them. ”*

- When a photographer of an image is identified, the photographer is either contacted or credited



# Copyright in the Archives

- Previously published items that are integral to a collection fall into “fair use” guidelines

HOWEVER

- Magazine and journal articles that are NOT from non-profit or scholarly associations should probably be left out of your digitizing project. Remember, these are available elsewhere!

# Questions?

## Contact information

- Mary Wallace
  - (313)577-4864
  - [maryjwallace@wayne.edu](mailto:maryjwallace@wayne.edu)
- Arjun Sabharwal
  - 419-530-4497
  - [arjun.sabharwal@utoledo.edu](mailto:arjun.sabharwal@utoledo.edu)



Digital Preservation: Starting and managing a digital preservation program by Mary Wallace (Wayne State University), Arjun Sabharwal (The University of Toledo) is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License](https://creativecommons.org/licenses/by-nc-nd/3.0/).