Application of Rapid Imaging Methods in a Library-Archive-Museum Environment

Brian Wilson – Benson Ford Research Center, The Henry Ford
brianw@thehenryford.org | twitter: @brilwil

1 Introduction

Digital camera imaging can produce access-quality images at several times the rate of flatbed scanner processes and dramatically increase the amount of digitized material derived from archival collections. The Benson Ford Research Center (BFRC) at The Henry Ford, Dearborn, Michigan, has launched just such a camera-based rapid imaging workflow aimed at supporting an organization-wide digitization initiative.

The BFRC Library-Archive-Museum (LAM) environment presents challenges to high volume imaging, however, including the blending of archival and museum metadata requirements; integration with existing 2D and 3D object imaging processes; and collection management system ingest and display of image data.

Development of and outcomes from the BFRC rapid imaging process, and approaches taken to meet LAM challenges are outlined.

Recto/Verso Capture

Rapid can translate to imaging of both front and back.

Wabash Ave. looking north, Chicago, ca. 1900; P.DPC.070163 [THF203264 – 265]

2 Need Something New

Initial imaging goal:
- 120,000 items over 2 years
- Metadata required for each image

Existing flatbed process:
- Single workstation
- 10 - 12 images per hour
- No metadata (done in separate process)
- At capacity with current work

New process must:
- Produce 300 - 400 images per day
- Produce metadata for each image
- Minimize equipment cost
- Minimize manpower cost

Material Size Range

0.75x1.5 to 22x30 inches

Matchbook
87.13.47.12 [THF204459]
2.25x1.5

Road Map
89.360.4 [THF205141] 22x30

3 Alternatives

Institutions using rapid imaging processes:
- Beinecke Rare Book & Manuscript Library*
- Yale University Art Gallery*
- Art Institute of Chicago*
- National Gallery of Art*
- San Diego Air & Space Museum

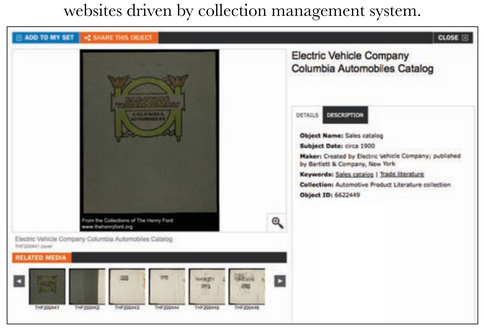
Additional information:
- OCLC Rapid Capture project led by Ricky Erway (http://www.oclc.org/research/activities/capture/default.htm)
- DIY book scanning sources (e.g. http://www.diybookscanner.org/)

Key points:
- Well-documented implementation at several institutions (e.g. http://beinecke.library.yale.edu/brbltda/dis/dishome.asp)
- Production rates in range required
- Images available for review on public Internet
- Costs competitive with flatbed scanner

(* See: Speed the Plow: Rapid Capture Digital Workflow session @ MCN2009. http://www.museumcomputernetwork.org/old-conferences/conferences/index.aspsubkey2524.htm)

Image Access

Images made available through public and internal facing websites driven by collection management system.



http://collections.thehenryford.org/

4 New BFRC Process

Key points:
- Access-quality output - Preserving source material
- Entire object captured - Not just content area
- Full color images with no grayscale conversion
- No post-capture color, skew, or object flaw corrections
- All sides of source captured - e.g. front and back of photographic prints
- Camera model also used by THF museum photo department

Equipment:
- Canon EOS 5D Mark II camera body
- 5616 x 3744 pixel sensor
- Canon 24-70mm and 50mm lenses
- Kaiser copy stand with daylight fluorescent lighting units
- Dell Optiplex 780 workstation (dedicated to imaging station)
- Software: Canon utilities; Adobe DNG, Bridge and Photoshop
- Costs:
  Camera and copy stand: $7,200
  PC hardware and software: $1,500

Equipment Setup

- U-shaped 7x8 ft. work area
- Black velvet covering for copy stand base - Provides neutral background and contrast for auto-cropping

Material types and sizes:
- Reflective materials only (to date)
- 0.75” x 1.5” up to 20” x 30”
- Loose and bound

Image specs:
- RAW Camera-out:
  400ppi min. for objects 14” x 9.4” or smaller; 25-30MB each
  Deleted after masters moved to network
- TIFF Masters:
  5616 x 3744 pixels @ 350ppi, uncompressed, 8 bits/channel, 50MB each average
  Stored in network preservation area
- JPEG Derivative:
  1200 x 1200 pixels @ 120ppi, 12 quality, 500-800kB each
  Uploaded to CMS for general access

Multipage Items

Imaging speed can also be applied by capturing all pages of multi-page items such as photograph albums.

1904 Vanderbilt Cup photograph album
Henry Austin Clark collection [THF203030 – 031]

5 LAM Issues

Higher scanning rate opens door to digitization of entire archival collections, not just select individual objects.

Object-focused selection, description, and access practices need to be adapted.

Selection
- Individual objects continue to be selected for exhibit, but now scanning entire folder or box

Description / Metadata
- Objects require individual catalog records, but required record metadata has been reduced
- Increased number of staff and departments involved in cataloging
- Object metadata cross-walked to DACS to facilitate archival collection cataloging

Workflow Integration
- 2D and 3D imaging using common data entry forms and conventions
- Cataloging and imaging to run as parallel not serial processes

Access
- Automating ingest of image files into collection management system to accommodate volume
- Revising internal and public-facing websites to display images in collection context

Object and Image File Metadata

- Object set based on VRA CCO minimum
- Image file set based on Dublin Core

Object ID	Color	Publish for Access:Intranet:Y/N
Object Type	Subject Term:Personal Name	Publish for Access:Internet:Y/N
Object Rating	Subject Term:Corporate Name	Publish for Access:Intranet:Destination
Current Location	Subject Term:Geographic Name	Publish for Access:Internet:Destination
Accession Lot	Subject Term:Topical	Image File – Title
Credit Line	Subject Term:Genre	Image File:Creator
Object Name	Subject Term:Events & Meetings	Image File:Source
Title	Subject Term:Local	Image File:Identifier
Collection Title	Measurement Type:Height	Image File:Creation Date
Creator Name	Measurement Type:Width	Image File:Imaging Tech
Place of Creation	Measurement Value:Height	Image File:Coverage
Creator Display	Measurement Value:Width	Image File:Rights
Date Created	Measurement Unit:Height	Image File:Publish for Access:Intranet:Y/N
Subject Date	Measurement Unit:Width	Image File:Publish for Access:Internet:Y/N
Materials	Task Data:Cataloger Name	Image File:Publish for Access:Intranet:Destination
Techniques	Task Data:Date	Image File:Publish for Access:Intranet:Destination

6 Production Rates

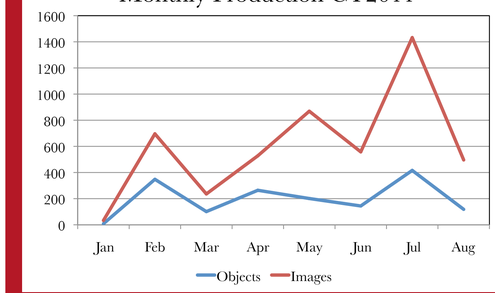
Imaging:
- Average: 45 images/hr or 8.5 objects/hr (mixed formats across multiple collections)
- Peak: 114 images/hr or 57 objects/hr (8x10 photographic prints)

Post-processing:
- Average: 50 images/hr (interneg, rotation, rename, exposure, crop, tiff conversion)

Object cataloging:
- Estimate: 4-6 objects/hr

CollMS image file ingest:
- Estimate: 20 images/hr

Monthly Production CY2011



8 Conclusion

Digital camera-based process achieves imaging rates far higher than existing flatbed process.

But, higher rate can expose other issues:
- Object-focused material selection
- Metadata creation process
- Amount of required metadata
- Catalog record QC and approval
- Online display system capabilities
- CollMS ingest speed

Imaging rate should include all phases of process:
- Selection, metadata, imaging, post-processing, and management system ingest.

Next Steps:
- Adapt process to handle transparent materials
- Further reduce metadata set
- Parallel path metadata and image creation
- Automate CollMS ingest
- Integrate images with finding aid content
- Create imaging station work instructions
- Train additional staff to use equipment