

## Lessons for Scanning, Part 1: Planning Your Digitization Project

Peter D. Verheyen  
Head, Preservation and Conservation  
Syracuse University Library  
September 24, 2008

[pdverhey@syr.edu](mailto:pdverhey@syr.edu)

## Introductions and Outcomes

- Who am I?
- What do we hope to learn?
  - How to select collections for digitization
  - Preparations and working on a project
  - Publishing collections
  - The impact of digitization on services

September 24, 2008 Peter D. Verheyen and Bradley Shipps 2

## Initial Considerations

- Type of Project
- State of Collections
- Infrastructure
- Staff Training
- Funding
- Standards

September 24, 2008 Peter D. Verheyen and Bradley Shipps 3

## Type of Project

- Can be dictated by state of collections
- Online Inventories (EAD, ...)
  - If not fully processed – processing project.
- Imaging or full-text
  - Item level access required

September 24, 2008 Peter D. Verheyen and Bradley Shipps 4

## State of Collections

- Are collections fully accessible at the item level?
- What is physical condition of items
- Copyright status

September 24, 2008 Peter D. Verheyen and Bradley Shipps 5

## Infrastructure

- Production
  - In-House
  - Outsourced
- Publication
  - In-House
  - Outsourced
    - OCLC's CONTENTdm through NNYLN
    - Flickr...

September 24, 2008 Peter D. Verheyen and Bradley Shipps 6

## Staff Training

---

- Experience with technology
  - Can dictate whether project completed in-house or outsourced
- "Professional" vs. Volunteers
  - "Standards" and supervision

September 24, 2008

Peter D. Verheyen and Bradley Shipps

7

## Funding

---

- Organizational funds
- Grants
  - Cost-share
  - Equipment
  - Salaries
  - Out-sourcing

September 24, 2008

Peter D. Verheyen and Bradley Shipps

8

## Standards for Master Files

---

- Have been established for most media and become routine.
- Master image should be use-neutral

September 24, 2008

Peter D. Verheyen and Bradley Shipps

9

## Let's Get Started

---

- Identifying the collection
  - Why this?
  - Current use?
  - Access?
  - Copyright status?
  
- Should always ask these questions

September 24, 2008

Peter D. Verheyen and Bradley Shipps

10

## Plan of Work

---

- Who needs to be involved?
  - Make sure EVERYONE is onboard. Who does what, when?
  - Tendency to inflate involvement.
- How much of their time is involved?
  - Usually expressed in % of effort.
- No experience with previous projects?
  - Run a test batch for in-house functions.
- Can project be completed in time?

September 24, 2008

Peter D. Verheyen and Bradley Shipps

11

## How are you want to pay for it?

---

- Funding Sources
  - Internal funds?
  - Grants? RBDB great, but NYS funds being cut.
  - Both? (Cost-share)
    - Many grants require institutional commitments. 50% not unusual.

September 24, 2008

Peter D. Verheyen and Bradley Shipps

12

## Budgeting

- Need to account for everything.
  - Staff (who, what level, fringe rate, ...)
  - Equipment needs
    - For production
    - For presentation
    - Long term planning.  
Digital = Buy Now, Pay Forever
  - Indirect costs
    - Required by some to cover admin costs, lights, ...

September 24, 2008 Peter D. Verheyen and Bradley Shipps 13

## So, you got a grant.

- When does the check arrive?
- Can you start early?
  - "Forward-funding"
- Does the plan of work still hold?
  - Staff changes
  - Budgetary changes
- Infrastructure
  - Are you ready?

September 24, 2008 Peter D. Verheyen and Bradley Shipps 14

## Workflows

- Make sure collection ready to go
  - Do not wait for the grant to start if there are things that need to be done before you get to the funded part.
- Start as early as you can
  - Things have a way of happening.
  - Hire staff and purchase equipment
  - Get training if needed

September 24, 2008 Peter D. Verheyen and Bradley Shipps 15

## Workflows

- Does everything need to be done in a linear fashion or can they be done on parallel tracks? E.g.:
  - Scanning
  - Metadata creation
  - Quality control
    - Should be ongoing
    - Do again before publishing, but be prepared for later fixes/tweaks

September 24, 2008 Peter D. Verheyen and Bradley Shipps 16

## Digitization – Organization

- Naming conventions
  - Unique within project
    - Directory structure
  - Unique within site
  - KIS(S) – Keep it simple.
    - Numeric – 0001.tif, 0002.tif, ...
      - Can also be descriptive if matches images in inventory and on object.
    - Descriptive – jhnoyes1.tif, jhnoyes2.tif...
      - Can get very confusing.

September 24, 2008 Peter D. Verheyen and Bradley Shipps 17

## Digitization – Master Images

- Should be use-neutral
- Tiff for images (or JPG2K)
  - 600dpi, RGB for most items.
    - Color helps capture tonal qualities even in b/w.
    - Grayscale for black and white negatives.
  - Lossless compression Guidelines freely available online.
- BACK UP, BACK UP, BACK UP
  - External hard-disk(s) are cheap.
  - CD/DVD media not "archival"

September 24, 2008 Peter D. Verheyen and Bradley Shipps 18

## Digitization – Screen Images

- In general, systems like CONTENTdm can convert .tif to .jpg for display
- Scale to fill most common screen dimensions (1024 x 768)
  - Remember to take browser/interface into account.
  - Compression of image for faster download
  - Smaller image less attractive for “misuse”

September 24, 2008

Peter D. Verheyen and Bradley Shipps

19

## Digitization – Rights Mgmt

- If it's online expect it to be “misused”
  - Provide good quality image, but make impractical to use in publications...
  - Smaller size, discrete watermarking, ...
- Provide hi-res images on demand
  - Have form for photo-duplication that states terms of use...
  - Important source of revenue (\$13K@SU)
- Advertise avail. of hi-res images.

September 24, 2008

Peter D. Verheyen and Bradley Shipps

20

## Metadata

- Vital to all digital projects
- Standards
  - MARC, DC, VRA, EAD, ...
  - Format can be adapted to other environments
    - Databases
    - Static pages
    - Migration
      - Create only once if possible.
- More details in next workshop, 10/8.

September 24, 2008

Peter D. Verheyen and Bradley Shipps

21

## Working with Metadata

- Work in “off-line” database or spreadsheet
  - Database has “prettier” interface
  - Can be exported to spreadsheet
  - Excel is your friend, really
    - Allows data to be manipulated easily
    - Especially practical for larger projects

September 24, 2008

Peter D. Verheyen and Bradley Shipps

22

## Publishing Options

- Own server
  - Maintenance – You're responsible
  - Database driven environment
  - Static pages
    - Maintenance unwieldy
    - Longer load times
    - Difficult to search
  - BUT
    - Often best/only option for small operations

September 24, 2008

Peter D. Verheyen and Bradley Shipps

23

## Publishing Options

- Hosted service
  - Maintenance – Someone else's problem
  - CONTENTdm
  - Facebook, Flickr, ...
    - May be practical, but beware of “terms of use” and YOUR rights.
    - Proprietary sites – What about long term preservation, extracting, records.

September 24, 2008

Peter D. Verheyen and Bradley Shipps

24

## Slide 21

---

L1 Covered in more detail on 6/6 in Metadata session.  
Library, 8/29/2007

## Impact on Public Services

- Be careful what you wish for!
- Increased digital and web presence = increased need of more staff to handle flood of questions/duplication orders
- Once your collections are online, they will be found by a global audience.
  - Indexing via Google
  - Have a plan for responding to requests for information and copies
    - Information on duplication costs
    - Information on managing your rights
      - Request to publish, fee structure, ...

September 24, 2008 Peter D. Verheyen and Bradley Shipps 25

## Impact on Public Services

- Increase of access to material makes for easier transaction?
- Yes and No.
  - YES
    - More direct ref questions (i.e. exact description and box #)
    - Increase in income for photo-duplication orders, other fees
  - NO
    - More questions = more staff time
    - More remote questions/orders = less visiting researchers

September 24, 2008 Peter D. Verheyen and Bradley Shipps 26

## Success stories @ SUL

- Saw increase in questions pertaining to/publication permissions for digital images.
- Some examples:
  - [Eisenmann](#) - Discovery Health Channel's *Sideshow Stars: The True Story*



September 24, 2008 Peter D. Verheyen and Bradley Shipps 27

## Biggest Impact - Inventories

- Mounting of collection inventories has the most significant impact on accessibility and use of collections
  - Allows researchers to see scope of collections
  - Assists staff because researchers are better prepared
  - Identifies good candidates for digitization

September 24, 2008 Peter D. Verheyen and Bradley Shipps 28

## Thank you

- For more information, please feel free to contact me at:
  - Peter Verheyen  
Head, Preservation and Conservation  
Syracuse University Library  
[pdverhey@syr.edu](mailto:pdverhey@syr.edu)  
315.443.9756

September 24, 2008 Peter D. Verheyen and Bradley Shipps 29