**Session:** Image Digitization

**Speaker:** Tom Rieger, Conversion Support Services Manager, Library of Congress

**Date:** Monday, 9/14, 11:00am-3:15pm

Tom's job:
- Planning digitization projects
- Oversees 14 different digitization units in library
- Make consistent across many different units, projects
- Know what requirements are

Need vs. effort
- Is it justified?
- Come up with a plan for the project - decide if it makes sense to spend the resources.

Digitization planning
Answer a few questions:
- Define Scope and Requirement
  - Identify what you have, what is important TO YOU
  - Digitization Survey
- How much time do you have to do this project?
  - Do you have the resources to get it done?
- What kinds of material do you have
  - each material requires a different workflow, equipment, skills
- What are your expectations? What is the end goal
  - Archival quality surrogate
  - Access copy, not so high quality
  - You may want to use high quality material in the future.

Tom is here to help! He can share information and advice!
Library of Congress is a resource, for the nation

How large is your project?
- Many times, unknown, BIG problem
- Can take months to figure out

What is the timing of the project?
- Event to present material, anniversary, grant funding
- How are you going to get your staff/resources/money coordinated
- Time you have left before materials degrade

Project Management at LoC
1. Request from service unit to digitize, information about the project - ask lots of questions up front. Team members:
   a. Conservation (assess materials)
      i. DO NO HARM - more important to preserve original than to digitize (digitize in the RIGHT way)
   b. Presentation and Access (how materials will be presented/used)
   . Do any access restrictions exist?
i. What is the idea of how the materials will be shared
   c. Back end programmer
   . The way things are made accessible
   d. Metadata
   . Do we have information, does it make sense
   e. Evaluate whether the project makes sense
   f. Contextual materials
   . What is the full story, what are all the pieces that make the package?

2. Analyze whether project makes sense to carry out
a. are there challenges in digitizing?
b. Decide priority based on materials, needs
c. Plan out before you start

Project 1: Make LoC collections available to the world

Project Plan
Good reason to have them - once you understand materials, you need to figure out how it is going to work.

Project plan sets your goals
- The best way to stay on target, have something to measure against
- Estimate how much can or should be done
- Meet every week to keep track of project
- Adjust plan if needed (extend deadline, add workers, increase hours)
- Take all preservation issues into consideration, add that into time (care of materials is part of plan)

Equipment
- Test system - overhead camera (attainable quality and price - a few thousand dollars)
- FADGI guidelines - 3 star images
- Creative solutions, can be done, example: Should not be used in archival collections, but the iPhone camera is pretty good!
- State of imaging today is so good, no excuse not to do this.
- Idea is to capture the ORIGINAL image, most accurately as possible

Flatbed scanner
- Does not recommend as a tool
- Overhead cameras are better
  - Can't squish a book in a flatbed
  - Get shadows on wrinkled pages
  - Flatbeds autocorrect - features are sometimes difficult to turn off
  - Conservation issues with putting material on glass and closing lid
  - Sharpening - autocorrect with scanners
  - Speed - many more steps in the process with a flatbed

Comparing results between flatbed and camera
- Tonality
- Mechanical interaction, scanners wear out

Question: Editing
- Keep master files
- Not RAW - they are more likely to become obsolete faster
- 16 bit tiff format
- All other files are derivatives

Migrating formats
- Be aware of how much material you are looking to store

Question maps
- scan or photograph small pieces and piece together
- overhead scanner
- Master is the individual segments
- Feed scanner - dangerous for old maps, put in mylar first, then put through

Cameras Tom recommends
- Lens is as important as the camera
  - Flat field lens - Macro, close focus
  - 50mm F2 compact macro $200
- Nikon D 8-10
- Canon 5Ds, 5D mark III
- Canon Rebel t1I - low budget

Part 2 hands-on

Different types of photographs, papers, drawings,

Taren Ober - image specialist
Geography and maps

Body: Canon 1D Mark III
A good lens - 85mm
Will cover whole field of capture
Lights

Canon software - found serial number, Canon download
Tethered camera - plugged in, live view of
Change camera settings from software
Saves, batch processing

DICE target to measure quality of image - next year should be available as an open source tool
OpenDICE
scan target - measures the performance of the device
Measuring grayscale, color, resolution through slant lines
    Many ways of measuring all the different levels
Pass/fail

*http://blogs.loc.gov/loc/2015/08/dice-y-digitization/

Be extremely level

ISO - film speed and sensitivity level

FADGI - better drafts coming October 1
http://www.digitizationguidelines.gov/

Photoshop and bridge
GIMP - Photoshop equivalent
IrfanView

Monitor calibration - reference calibrator
Buy a "good" monitor - percentage of adobe 1998 colorspace, 98-99%
Self-calibrating monitors - 2,000
http://www.colour-science.com/

Turn off auto-corrections

Example - engraving - resolution problem to display lots of small lines
Example - map - sensor reads the map differently than the eye
Example - Big Boy photo, image not captured correctly, color bar

Color bars - measuring white, black, grey measuring whether values change from object to object

Environment for imaging
- follow fadgi specs
- lights
- grey walls
- neutral

Printing Images
- Selling reproduction prints
- Inkjet printing is not quality, laser printing started in early 2000s
- Before 2000

Tintype image capture demo
- Make sure item in straight
- Choose settings - or leave on auto
- Instead of RAW file...
- -Save master image - save (TIFF) (file size sometimes can be difficult, ex. of maps as jpeg2000)
- -Then create derivatives - access copies (lower quality, smaller size, easier to access) (jpeg)
  - o -Sharpening - enhance contrast between one and another edge - do it too much and it looks like everything is outlined (halos)
  - o -Different materials call for different sharpening

**JPEG VS JPEG2000**
jpeg: raster image
jpeg2000: mathematical, coding the image
Jpeg is not an archival format
   -possible degradation

**Question**
Have raw and jpeg
Can export from raw file to tiff (Camera Raw option in PS), and batch the process

**Question**
Film scanning with a camera

Find old photographic enlargers - film gates
Light source - light box (daylight corrected)
Macro lens

**Question - Maps**
Stitching images together - big maps
Take into photoshop, place next to each other
Photomerge in photoshop - but also keep the original individual files
Autopanogiga - Kolor

NEDCC - photoshop first
3D imaging - when things are not flat

Imaging technology improving very quickly, getting very cheap

Color
Color Profiling - ICC color management
DICE measures
    within 3 deltas (1 delta = observable difference seen)
Example - 13 deltas off

White balance important - easy to tell

When exporting - there is an option to save as 8bit vs 16bit

Exposure or contrast
Gamma - decision of how to interpret the change between white and blacks (interpreting midtones)

Uniformity - making sure things are the same across entire image
tilt of camera,

OpenDICE
Attempting to get features  - and additional features
-geometric accuracy
-skew
-how clean is your imaging

Glossary of terms
http://www.digitizationguidelines.gov/glossary.php

DATE: Tuesday, 9/15, 11:00am-3:15pm

**Image Digitization**
- Conversion Support Services Manager at LOC
- Provides support and guidance to all library digitization efforts
- Figures out how to do things better!
- Also works on LOC and general digitization standards with FADGI
  http://www.digitizationguidelines.gov/
- Co-ordinates Still Image Digitization Workroom
- New standards coming soon

This hour is about best practices

**Digitization Planning**
- Scale makes a difference, you might get away with minimal planning for a small collection, but imaging working at LOC - requires much more robust planning

4 Questions
- What's the scope?
  - All documents? photos? film? audio? mixed?
  - Scope will inform your equipment needs, setup and process
  - If you plan effectively, you can do almost anything with a given budget
- What's your timeline?
  - Specific event or anniversary coming up? (this is quite common)
  - Is your proposal possible in the time allotted?
- What kinds of materials are involved?
  - 18th century manuscript? 20th century paper? Photos of all types?
  - All need different handling, may require conservation before they can be handled or digitized
  - ***The originals are FAR MORE important than their digital forms, do NOT digitize if it will damage the materials
    - Ex: 18th century document that was stored folded - do not unfold it, it is probably brittle!
    - Ex: staples and fasteners, we tend not to remove parts, treat things as artifacts
    - Assess all this before you start doing the work, it is essential pre-planning
  - Build those considerations into your project plan
  - Preparation, Art handling, Collections Care Information network
    http://www.paccin.org

- What's your goal or purpose?
    - This will help determine your equipment needs
    - Decent equipment can be had for quite reasonable costs

**Have a project Plan**
- Are we on track?
- What are our blocks?
- Include a timeline, and daily/weekly/etc… expectations so you can compare against them at set intervals
- If going too slow, maybe you've chosen the wrong equipment?
    - Ex: flatbed scanner - used far too often in projects
    - They are very slow
- There are resources available for helping to pick the right tools
    - Even mid-sized colleges are great options to work with - usually decent equipment and cheap student labour

Set your production levels
- Sprint or marathon?
- What pace are you working at?
- It is realistic?
- Is it putting your materials at risk?
- Can be difficult to balance speed with care

**Equipment**
Three elements involved in selection
- Speed of operation
- Cost of device
- Quality of the results
The classic "pick any two" conundrum
Consider working off-network, on a dedicated digitization system

**Storage**
SSD (Solid State Drive) vs HD (Spinning disk)
- HD - spinning platters writing magnetic data
- SSD - newer and faster
    - Memory chip, no moving parts
    - Overall better for moving files around
- For near-term storage (routine access, but not your working device)
    - Use a RAID system (redundant array of independent disks)
    - Series of redundant hard drives with a safety protocol built in
- For long-term storage
    - The only acceptable technology currently available is LTO (Linear Tape Open 6) Tape
    - It is an old technology (WW2)

- o Lasts 15-30 years, but is updated every few years and only can access 2-generation old tapes with current generation of devices
  - o Thus, regular migration is required (3 years, 5 at the max), very very large task for archives
  - o 6 to 6.5 TB per tape (current generation)
  - o And of course, you cannot only have one copy of that LTO tape
    - Minimum 3 copies!
    - 3-2-1 rule - 3 copies, 2 of which are portable, 1 of which is someplace else
    - 3 copies on tape is fine, because file sizes (each tape being 6 TB) kind of make using another format tricky
    - Would be soooo good to have a non-magnetic long-term storage media
  - o LOC cycles tapes every year, discards them every 2 years
    - Have a schedule, stick to it
- Costs of everything keep coming down
- FADGI 3-star - minimal acceptable quality used at LOC, and most common
  - o 4-star is super high quality, and pretty unnecessary
- Commonly using 20MP cameras in high quality digitization centres
  - o Not more than a few hundred dollars for a body and that again for a lens
  - o But good software plays a role

**Flatbeds**
- Not recommended tools for archivists
- Don't actually produce really really good images when compared to DICE (Digital Image Conformance Evaluation) program
  - o Human eyes cannot judge the quality of a digitized image
  - o LOC/FADGI developed testing program to evaluate quality of digitized images (not born digital)
  - o Also a proprietary version known as [GoldenThread](#) (but very pricey)
  - o Just finished recoding DICE as open-license, coming out early 2016 hopefully
- Terrible for books! You WILL damage them
- Also not good for film
- Definitely not for Special Collections, you will damage fragile materials
- Can't do big original items

Everything at LOC uses contact-less imaging, taken face up, from above

**Planetary Scanners** have a line-element sensor that builds the image one row at a time
- Can be high quality
- But can have colour issues

**Cameras** images their entire field of view all at once

**Photographs**
- Look and feel of imaging from different devices is notable
    - Cameras give a smoother rendition than flatbed scanners - just looks nicer
    - LOC almost only uses cameras

**FADGI**
- Federal effort to measure quality of images across country and institutions

**Digital Imaging**
With Taren Ober - Digital Imaging Specialist

Working with a 20MP camera
- Can get a modern equivalent for about $500
- That's about the minimum, but is definitely good enough

Looking at a range of different image types today
- All reflective media (ie: not film)
- Book page, photo, print, letter, etc..
- When handling photos use gloves (because of the oil on your fingers), but not when handling paper (gloves make you clumsy)
- We have pre-digitized samples of everything to save time
    - Have a camera and scanner version of each for comparison on thumb drives
    - Each drive has 3 folders - flatbed, camera, combined
        - Flatbed has as much auto software turned off
        - Ditto for camera, which is how you want to work

**Dealing with unknowns**
- We turn to standards
- Object level target (smaller target)
    - Provides a fixed, unchanging reference point
    - Example is DICE/golden threads target
    - Other example is the device level target (larger target)
        - 300-400 ppi (points per inch)
        - used to help identify resolution, image density, tone scale
    - Targets are key to a professional digitization program
        - Especially for cultural heritage, where accurate representation is really important

**Book Page Imaging**
- Flatbed scanner not good for collections materials, going to do some damage
- Use a copy stand so you don't damage or stress the materials

- o We're not using any extra software, came stock with the camera (Canon)
- o Offers a live-view feed through the camera
- o We want to have everything set up as straight as possible
- o Can use full manual, full auto…. tailor to your needs
- o Aim to use a lower ISO - less noise in image, dependent on available light
- o Too much ambient light in this room, aim to use a darker room, need better lights as well (new LEDs are good, strobe is best, aim for daylight balanced lights)
- o Stand should be sturdy and level, beyond that, not super important
- o Camera should be level

**Files**
- FADGI 3-star, aim for 300ppi minimum for most materials
- PPI (resolution) is more important than file size
- Be consistent, keep 300ppi (or whatever) for all your images on a given project
- Camera height above scanner will generally correspond to a given ppi
  - o Be able to mark the camera height for consistency so you're not zooming and adjusting
- OCR (Optical Character Recognition) works best at particular settings (300 or 400 ppi)
  - o Figure out what your software uses and leave it there
- Geography and Maps (LOC) images at 300ppi and in JPEG2000 due to space constraints
  - o Everyone else at LOC uses uncompressed TIFF
- Negatives and film will need higher resolution, to capture film grain
  - o eg: 8x10 glass negative, 1000ppi
- Capture images in RAW format - pretty common now
  - o You can edit formats after the fact, just make sure your master is something like RAW
  - o RAW is proprietary to each camera brand
    - ▪ Don't use them as your archival master!
    - ▪ Proprietary file formats are not sustainable
    - ▪ Only a fraction of the image file formats developed are still in use
    - ▪ Use TIFF as archival…. at least for now
    - ▪ Save as 16-bit (ie: 64000 colours), captures many more colours than 8-bit (only 256 colours)
    - ▪ Proprietary files might not be supported from generation to generation
- Question: what software to use? Especially since photoshop is now all moving to Creative Cloud type services
  - o Big issue when online updates are required (especially for LOC)
  - o Try to use older, well established software
  - o Consider using GIMP (free alternative to photoshop) http://www.gimp.org/
  - o or IrfanView http://www.irfanview.com/

**Hands-on**
- Imaging a tin-type
- Old, kind of lumpy, actually on tin (or probably iron or steel, actual tin-types are rare)
- We aim capture things as they appear, not enhanced!
  - Often asked to image back of an object, there might be hidden information there
- Try to carry along edges, not touching the image/colour patches
  - Also important for handling targets
- Place target near, but not touching the item
- Align everything nice and straight, but 90 or 180degree rotations aren't an issue
- Using automatic, not manual, camera shutter settings
- For white balance
  - http://www.rmimaging.com/products.html
  - https://www.phaseone.com/en/Imaging-Software/Capture-One.aspx?
- All lenses will produce some distortion, due to their shape (called fall off)
  - we can compensate with software (see rmimaging, above)

What is digital ISO?
  - Conceptually, works the same as film (higher ISO = more sensitivity to light)
  - And higher ISO = more noise  (film grain, digital noise)
- What about light-sensitive objects? Don't want to expose to too much light?
  - Might have to use a higher ISO to reduce exposure time
  - BUT, don't use continuous exposure! Use a strobe flash instead

Imaging a regular printed item
- Need to adjust camera height for image size, then refocus for adjustment
- File shot in RAW, but can do some adjusting after
  - In Adobe Camera RAW, or other software
  - Adjust daylight balance, temperature, exposure

  - Currently using stock Canon software
  - Can see exposure histogram
  - All digital imaging devices sharpen, its unavoidable, but you want to use as little as possible. Too much inserts irremovable artifacts
    - Better to have a slightly soft image, than too sharp and crunchy
  - Histogram: visual representation of every tone between white and black in the image
    - Gives you an overview of the image, helps avoid clipping off high or low tones - keep everything within the graph scale
    - There's no ideal shape - an image that is very dark will have more at the bottom, etc….

Why RAW?
- Offers more capabilities later on
- As close to unformatted, unstructured data that comes off the camera sensor as possible
- Relatively free of colour balancing, exposure controls, etc… from the camera
- But there is no universal RAW format, so not great for archival use
  - Closest to archival is adobe DNG (Digital Negative) format, but not in use professionally yet
- Ex: photographer wanted to donate his collection all as RAW files
  - But then how do you display to the public, or store long term?

DICE
- Digital Imaging Conformance Environment
- International ISO standard for measuring resolution
- Works with LAB colour matrix
- Looking at tone (light-dark), colour (hue), and resolution (center to edge)
- Hopefully a version of open source DICE is coming soon
  - But the targets might still be pretty expensive
- ICC (International Colour Consortium) colour profiling
  - If you shoot an image without an ICC profile, you don't really know what the camera's doing
  - But manufacturers tune their devices to a specific ICC profile, which takes the incoming RAW file, and auto-adjusts to correct the input from the imaging sensor
  - We can very precisely measure colour, and get a reference point for aligning ICC profiles
  - So yes, when using DICE, use the embedded ICC profile when prompted
  - But, which profile do you use?
    - RGB? or one of a million other options?
    - SRGB or Adobe1998 - those two are the US accepted standards
    - SRGB more or less represents TV set colours
      - Developed because most TV sets could only display a smaller range of colours - why produce colours that couldn't be displayed?
    - Adobe1998 is the safer bet, because it has a wider range, especially if working with 16bit colour
    - Bigger space would be nicer, but is not especially practical at the moment
    - Exceptions might be things like newspapers with a very narrow colour profile
    - Also now a lot of web browsers are colour-managing, and showing Adobe1998, not just SRGB
- Open target > "crop and identify"
  - Lets you re-align your imaged target if needed

- Gives you a summary, if you passed the conformance test or not
  - Tone Scale - range of tones from white to black
    - Natural progression following Gamma 2.2 - a well-accepted description of that shift, based on how we see colour
    - Might need to adjust lighting, or gamma
- You generally trust your eyes, but that is looking at a screen, in a poorly lit room
  - We would never actually image in a room like this (better would be something dimmer, with a brightness like 50 lux)
  - Standard reference image from to calibrate/balance/set colours
    - http://www.colour-science.com/
  - Put money into colour set-up
    - You can get a decent camera for cheap, but then blow it all with bad colour settings
- Delta from A
  - Delta - one visually identifiable change
- There is also a standard colour checker target, and you can use that to build the cameras colour profile

  - Profile your camera with this target (maybe every few years)
  - Then it can probably actually pass the DICE target test (do it every time you turn the camera on - do it all the time)
    - www.delt.ae is a free, online checker - almost as good as DICE
  - Then use the image target in every shot, use that to calibrate each image