Digitization REVEALED: 101 Crash Course on the Fundamentals of Digitizing Archival Collections from Start to Finish

Marina Georgieva

University of Nevada, Las Vegas, marina.georgieva@unlv.edu

Follow this and additional works at: https://digitalscholarship.unlv.edu/libfacpresentation

Part of the Archival Science Commons, and the Collection Development and Management Commons

Repository Citation


Available at: https://digitalscholarship.unlv.edu/libfacpresentation/163
DIGITIZATION REVEALED

101 crash course on the fundamentals of digitizing archival collections from start to finish

MARINA GEORGIEVA

Nevada Library Association Annual Conference
October 14, 2018
Las Vegas, NV
ABOUT ME

- Graduated from University of Wisconsin – Milwaukee
- MLIS with concentration in Information technologies
  - Digital libraries
  - Graphic and Web design
  - Metadata
  - Information architecture and UX design
- Visiting Digital Collections Librarian at University of Nevada – Las Vegas
- Professional passion
  - Project management
  - Digitization
  - Metadata design and management

Digitizing archival manuscripts on Phase One camera (2018)

More about my work at www.marina-expertise.com
TOPICS

- Collection Inventory and Assessment
- Funding
- Digitization technology
- Team
- Workflow
- Platform
- Standards and sustainability
COLLECTION INVENTORY AND ASSESSMENT

WHY DIGITIZE?
• For access
• For preservation
• On demand
• For specific occasion/exhibit

ANALYZE THE COLLECTION
• Priority list
• Research value
• Collection size
• Collection level of processing (Finding Aid)
• Reuse of existing Finding Aid
• Condition of materials
• Format of materials
• Has it been digitized as part of another project?

ASSESSMENT
• Large scale or boutique approach?
• Digitize all or selected materials?
• Who does selection?
• In-house capacities
• Outsourcing options
FUNDING

1  Funding types

2  Grant-funding opportunities
INTERNAL FUNDING

- Types of digitization projects
  - Reducing archival backlog
  - Preservation
  - Access (esp. high demand collections)
  - Scan on demand (for class or event)
- Staffing
  - Existing staff
  - Students assistants
  - Interns
  - Volunteers

GRANT FUNDING

- Types of digitization projects
  - Specific project that fits the grant scope
  - High priority collection that requires more resources and is eligible for grant
  - High research value collections
  - Collections in desperate need for preservation
- Staffing
  - Specially hired and trained project staff
  - Specially trained students assistants
FUNDING

Grant-funding opportunities

National Endowment for the Humanities
https://www.neh.gov/grants/listing?keywords=-digitization

Library Services and Technology Act
https://nsla.libguides.com/2018LSTA

Institute of Museum and Library Services
https://www.imls.gov/grants/apply-grant/available-grants

National Historical Publications & Records Commission
https://www.archives.gov/nhprc/apply/eligibility.html
Types of digitization technologies

Selecting the best fit
Types of digitization technologies

1. Flatbed scanner
2. Book scanner
3. Camera systems
4. Large format scanner
5. Transparencies scanner
6. Microfilm scanner
DIGITIZATION TECHNOLOGY

2

Selecting the best fit

CONSIDER THE COLLECTION

- Format of archival materials
  - Microfilms
  - Transparencies
  - Reflective materials
  - Oversized materials
  - Manuscripts
  - Books

CONSIDER THE PROJECT

- Scale
- Scope
- Timeline
- Team

CONSIDER THE TECHNOLOGY

- Equipment you already have
- Equipment you need to purchase
- Equipment you need to outsource
**TEAM 1**

**Assessment | decision-making**

**PROJECT ASSESSMENT**
- Grant-funded vs. internally funded
- Project priority
- Project deadlines
- Project specifics
  - Large-scale or small
  - Scan on demand
  - Target audience

**DECISION-MAKING**
- Team size
- Roles (positions)
- Part-time vs. full-time

**HIRING**
- Internal staff vs. external hires
- Experienced vs. non-trained
- Professional staff vs. student assistants
ASSIGNING ROLES

- Train narrow specialists for particular tasks vs. train people universally
- Advantages and disadvantages
- Main roles in digitization:
  - Workflow manager
  - Staff manager
  - Metadata manager
  - Metadata creators
  - Digitization specialists

TRAINING

- New hires training
- Refreshers
- On-going training
- Training documentation
- Self-training guidelines
WORKFLOW

1. Preparation of physical materials
2. Digitization of archival materials
3. Image processing
4. Batch import
5. Metadata (object description)
6. Quality review
**STATUS CONTROL FILE**

- Comprehensive list of all collections in a project
- Provides vital information for tracking progress
  - Digital ID
  - Collection name
  - Workflow segments
  - Initials and completion dates

**MASTERFILE**

- Comprehensive list of all items in a batch
- Fields are replica of collection MAP
- Used for data import in DAMS
- Records initial descriptive and technical metadata
  - Digital IDs
  - Descriptions
  - Technology
  - Formats

**DIGITAL OBJECTS**

- Compound objects
- Single objects
- Not necessarily replica of the structure of the archival folder

**ARCHIVAL OBJECTS**

- A.k.a physical objects
- Described in Finding Aids
  - Box level
  - Folder level
  - Item level
Some definitions

### Status Control File

<table>
<thead>
<tr>
<th>Digital ID</th>
<th>Collection Name</th>
<th>Initials</th>
<th>Date</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ph141001</td>
<td>Imperial Palace</td>
<td>n</td>
<td>1/1/2012</td>
<td>Item 1</td>
</tr>
<tr>
<td>ph141002</td>
<td>Imperial Palace</td>
<td>n</td>
<td>1/1/2012</td>
<td>Item 2</td>
</tr>
<tr>
<td>ph141003</td>
<td>Imperial Palace</td>
<td>n</td>
<td>1/1/2012</td>
<td>Item 3</td>
</tr>
</tbody>
</table>

**MasterFile**

- A detailed spreadsheet showing data related to MasterFile. The spreadsheet contains columns for various metadata and statuses, with specific details for each entry.

**Workflow Diagram**

A workflow diagram illustrating the process of managing and controlling files, with specific sections for scanning, metadata, and QC processes. The diagram includes icons for file control, status tracking, and workflow progression.
WORKFLOW

1 Preparation of physical materials

DOCUMENTATION
- Analyze Finding Aid
- Prepare MasterFile
- Object type
  - single objects
  - compound objects
  - textual
  - visual
- Object format
  - negatives
  - reflective materials
  - 3D objects

OBTAINING COLLECTION
- Finding Aid
- Get collection
- Collection inventory
  - Does Finding Aid description correspond to what’s in boxes?
  - Is Finding Aid description on folder level or on item level?

PROCESSING COLLECTION
- Condition of materials
- Documenting objects
- Folder vs item level decision
- Grouping for efficient digitization
- Updating documentation
WORKFLOW

2

Digitization of archival materials

DIGITIZATION PREP
• Technology selection
• Set up scanning sessions
• Group in batches if materials are heterogeneous
• Set up documentation
  • MasterFile
  • Status Control File
• Set up file naming

DIGITIZATION
• Scan!
• Update documentation simultaneously
• Keep track of items and batches
• Transcribe all peculiarities of the objects
WORKFLOW

3

Image processing

ACTIONS

• Straightening
• Cropping
• Color correction
• OCR (textual objects)

The sequence of actions and used software depends on the technology used for scanning

FILE NAMING

• Digital IDs are critical for object retrieval and file preservation
• Double-check file names
• Follow the collection naming conventions
• Organize compound objects
• Update documentation

EXPORTING

• Create a collection destination folder for temporary home of archival images
• Export TIFFs and JPGs
• Organize compound objects
WORKFLOW

4 Batch import

DOCUMENTATION

• Assign batch number
• Update MasterFile
• Update Status Control File
• Generate Tab-delimited .txt file for the import
• Get the path to the archival images

IMPORT

• Compound objects w/ parent level metadata
• Compound objects w/ children level metadata
• Single objects
• OCR’ed objects w/ .txt files
• Troubleshoot if necessary

ASSIGN

• Check the transcripts (for OCR’ed objects!)
• Communicate the batch number to the metadata creators
OVERVIEW

- Object description
- Promotes consistency across collections
- Enhances easy object retrieval
- Supports faceting
- Supports subject searching | browsing

METADATA GRANULARITY

- Rich vs. basic metadata
- Collections size – granularity differs across collections
  - Large-scale projects
  - Boutique collections
  - Digital exhibits
- Material format – some collections need less granularity
  - Newspaper collections

METADATA ROLES

- Administrator
- Managers
- Creators

Most important digitization segment from user perspective!
Metadata (object description)

**METADATA APPLICATION PROFILE**
- Fields
- Encoding scheme
- Occurrence
- Obligation
- Collection-specific vs. shared

**INDEXING GUIDELINES**
- Cheat sheet for metadata creators
- Clarifies MAP
- Disambiguates
- Rules and examples

**CONTROLLED VOCABULARIES**
- Local CVs
- Authority files
  - AAT | Art and Architecture Thesaurus
  - FAST | Faceted Application of Subject Terminology
  - LCSH | Library of Congress Subject Headings
  - LCNAF | Library of Congress Name Authorities
  - MESH | Medical Subject Headings
  - TGN | Getty Thesaurus of Geographic Names
  - ULAN | Union List of Artist Names
- Collection specific vs shared

*Most important digitization segment from user perspective!
WORKFLOW

Quality review

6

ACTIONs

• Image quality
• Metadata quality
• Transcripts
• OCR quality

FEEDBACK AND REVISIONS

• Communicate problems to metadata creators
• Revise and correct
• Keep list of typical mistakes for training purposes

GOAL

High quality metadata
= consistent collections
= linked data readiness
= increased usability of digital collections
= happy researchers!
## PLATFORM

### FRONT-END FEATURES
- Easy navigation
- User-friendly
- Customizable
- Responsive design
- Supports multiple collections
- Supports searching and browsing
- Supports advanced search
- Supports faceting

### BACK-END FEATURES
- Easy upload of data sets
- Easy export of data sets
- Easy to use by staff
- Intuitive administrator interface
- Supports multiple file formats
- Supports shared vocabularies across collections
- Cloud-based vs. server-based

### POPULAR DAMS
- CONTENTdm®
- Omeka
- Drupal™
- Veridian

---

The DAMS interface is your presence in the web! Keep it neat to provide outstanding user experience!
STANDARDS AND SUSTAINABILITY

STANDARDS

• Dublin Core Metadata Element Set (DC)
• Metadata Object Description Schema (MODS)
• Metadata Encoding and Transmission Standard (METS)
• eXtensible Markup Language (XML)
• Encoded Archival Standard (EAD)

SUSTAINABILITY

• Iterative approach
• Reuse, adapt, improve existing workflows
• Cross-collection controlled vocabularies
• Adopt best practices
• Have internal procedures and guidelines in place
THANK YOU!

QUESTIONS?

Visiting Digital Collections Librarian
University of Nevada - Las Vegas
University Libraries
4505 S Maryland Parkway
Box 457041
Las Vegas, NV 89154
Tel: 702-895-2310
marina.georgeiva@unlv.edu

www.marina-expertise.com