

DALNET Digitization Workshop

Ruth Ann Jones
Digital Projects Manager
Michigan State University
Libraries

DALNET

Overview

- Setting project goals
- Planning workflow
- Effective training

Agenda for the afternoon

• 1:15-2:15	Setting project goals
• 2:15-2:30	Questions
• 2:30-2:45	Break
• 2:45-3:45	Planning workflow, Effective training
• 3:45-4:00	Questions

Setting project goals: overview

- Choose collection to digitize
- Decide how to digitize
- Estimate time & cost requirements

Choose collection to digitize

- Patron interests
- Copyright issues
- Availability of originals
- Interests of potential funders
- Consistency of formats

More info: [Selecting Research Collections for Digitization \(CLIR\)](#)

Choose items to digitize

- Will individual items be damaged by handling?
- Digitize entire collection or a representative subset?
- How many items needed for a "critical mass"?

Deciding how to digitize: overview

- Formats
- Standards
- Accessibility

Formats

- Image scanning
 - Image scanning of images
 - Image scanning of text
 - More info: Cornell University's Digital Imaging Tutorial

Formats

- Text transcription
 - Civil War letter
 - Polk County Obituaries
 - More info: A Gentle Introduction to SGML (parent of XML)

Formats

- Optical Character Recognition

[Jan Koutnik's Java OCR demo](#)

[Original image](#)

[Scanned text](#)

[More info: Pacific Lutheran University "Basic Scanning" help pages](#)

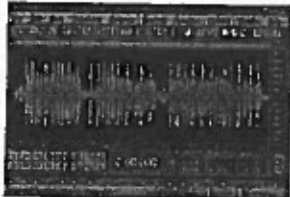
Markup for full-text searching

- XML (eXtensible Markup Language)

[Begin Now \(source code version\)](#)

[Begin Now \(display version\)](#)

Sound and video digitization



- [More info: Marr Sound Archive digitization demo](#)

Digitizing multidimensional objects

- Museum objects and artifacts can be seen from all sides and rotated by the user

Object for Feeding America: The Historic American Cookbook Project

Standards

- Use non-proprietary data formats
 - Text transcriptions in ASCII only
 - ASCII Character Chart
 - Image files in TIF, JPG, GIF

Standards

- Follow established technical standards for digitization quality
 - accuracy of text transcriptions
 - image resolution and color depth

Accessibility

- Preservation copies vs. use-copies
 - Bandwidth calculator
- Site searching
 - Search engines
 - Dynamic interfaces
 - SSB search screen SSB source code
- Resource discovery
 - Metadata for digital objects

Estimating time/cost requirements

- Flatbed image scanning: 15-25 images/hour
- Sheet-feed image scanning: 30+ images/hour
- Optical character recognition: 10-30 pp/hour
- Typing: 5-10 pages per hour, plus proofing
- Sound: Length of sound file plus cueing time

Why are these estimates so broad?

- Other factors affecting speed of digitization processes:
 - Fragility of material
 - Computer speed
 - OCR accuracy (time spent on corrections)
 - Processing of archival copies to produce use-copies
- No substitute for in-house time estimates!

More help with cost estimates

- Scanners
[CNET Buyer's Guide to Scanners](#)
- Optical Character Recognition Software
[CNET OCR Software Comparison Tests](#)
[CNET OCR Software Specifications Chart](#)
- Search engines
[SearchTools.com](#)
- XML editors
[XMLSoftware.com list of XML editors](#)

• Questions?

• 15-minute break

Planning Workflow: overview

- Devise workflow
- Develop specific procedures
- Gather feedback after work begins

Devise workflow

- Divide work into stages
Breakpoints: software changes; format conversions
- Consider time constraints
Deadlines
Hours per week staff are available
Hours per week equipment is available
Do you need to provide at least partial access ASAP?

Develop specific procedures

- Prepare by *doing the work*
- Mistake-proof the system as much as possible
- Build in accuracy checks
- Build in data backups
- Plan recordkeeping
- Plan file and directory organization
- Consider how problems will be handled

Gather feedback after work begins

- Talk with staff performing digitization
- Check work after training *and* after changes in procedure
- Keep detailed records of work completed *and* time required
Example: typing worksheet
Example: tracking worksheet

Effective Training: overview

- Prepare by *doing the work*
- Develop clear written instructions
- One-on-one training

Plan by doing the work

- Assists in gauging complexity of a task
- Ensures that trainer understands details of a task
- Facilitates problem-solving

Develop clear, written instructions

- Document the sequence of multi-step, mechanical processes. *Example.*
- Provide guidelines and examples for interpretive processes. *Example.*
- Create reference materials on directory structures, filenaming, etc. *Example*

One-on-one training for digitizers

- Explain how task relates to overall process
- Demonstrate task
- Observe first (few) repetition(s) of task
- Be easily available during first work session
- Encourage questions

More information

- Internet Public Library, *Pathfinder: Digitization*
- SUNY at Buffalo, *Resources for Digitization*
- University of Southern Mississippi, *Digitization Resources Online*

- Questions?
