

Streamlining Mass Digitization for Archives Materials

Joshua Ranger – Forrest R. Polk Library | University of Wisconsin Oshkosh

Challenge: Bring digitization costs-per-page for paper-based archival materials down sufficiently to exploit large collections.

Solution: Employ University of Wisconsin Digital Collections streamlined model to achieve reduction in costs. Improve browsing and searching functions to make model more acceptable to users.

UWDC Models: Two models were employed in experiment:

Control (traditional) model:

Full color scans, isolated documents, descriptive metadata includes date, location and names of correspondents. Uncorrected OCR. \$1.53 per page.

Streamlined (experimental) model:

Bi-tonal scans made from photocopies, documents NOT isolated from each other, NO descriptive metadata on individual documents. Uncorrected OCR. \$0.33 per page.

See them employed: <http://digicoll.library.wisc.edu/WI/subcollections/AJamesAbout.html>

USABILITY Testing: Seven graduate library students, seven undergraduate history majors. Participants undertook six research tasks employing different features of both the control and experimental models. Interviews and post test followed.

- On the whole, both groups preferred the control model to the experimental, finding it easier to use, navigate and search. Graduate students were more critical of the streamlined version, however they did not have prior archival research experience.
- Undergraduates reported they were more likely to use the experimental model if available for research in the future.
- Students in both groups reported wanting MORE not LESS metadata than even the control model offered, desiring a short abstract for each document.
- Most students prefer searching rather than browsing. Expect full text, Google-like search capability.

Possible Strategies:

- Modify CONTENTdm or create next generation product to allow for improved browsing function.
- Explore automating the isolation of documents.
- Investigate OCR options to create a better transcription.
- Consider user-contributed transcription opportunities.