

Rapid Capture in University Archives

A Model for Mass Digitization
of Institutional Content

Erik Moore
University of Minnesota
August 24, 2011

UNIVERSITY OF MINNESOTA
LIBRARIES

As more and more is born digital
and a new generation of users
grows up with digital as the default
mode of delivery, resources that
are not in digital form will be
'orphaned' over time because they
are in 'obsolete' formats.

Abby Smith, Council on Library and Information Resources
(ARL, Recognizing Digitization as a Preservation Reformatting
Method, 2004)

UNIVERSITY OF MINNESOTA
LIBRARIES

There's an illusion being created that all the world's knowledge is on the Web, but we haven't begun to glimpse what is out there in local archives and libraries.

Material that is not digitized risks being neglected as it would not have been in the past, virtually lost to the great majority of potential users.

Edward L. Ayers, University of Richmond (New York Times, "History, Digitized (and Abridged)," March 10, 2007)

UNIVERSITY OF MINNESOTA
LIBRARIES

Discoverability

- Description is not equal to discovery
- Users expect discovery & delivery to coincide
- Discordant to experience this disconnect in the library environment

UNIVERSITY OF MINNESOTA
LIBRARIES

Large-Scale Digitization

- Traditionally grant funded
- Image based, item level focus
- Eye candy & high profile collections
- Outreach for a specific audience

Sustainable Model

- NARA & OCLC: integrated within normal operations
- Best means to achieve large-scale conversion
- Digitize for access. More is better.

Material

- 20th century
- Mass produced
- Informational value
- Uncataloged hardcopy

Pilot Project

- Academic Health Center History Project
- Digitize & make available via IR
- Solicited bids \$0.69-0.90/scan
- Proposed in-house solution
- 80,000 1st year at \$0.07/scan
- Proof of concept

DIY Digitization

- Equipment less than \$6,400
- 2 sheet-feeding scanners
- Imbedded in workflow (student hours)
- Average 200,000 page scans per year
- Completed 700,000 scans
- Less than \$0.01 per page

Selection

- Serial publications & records
- Material now collected digitally
- Collection or series level, not eye candy

Delivery

- Institutional Repository - University Digital Conservancy (conservancy.umn.edu)
- Access IR interface & search engines
- RSS feed (Reader, RefWorks, etc.)
- Not the sole distributor of our content
- Discovery happens elsewhere

Deliverables

- Downloadable PDF (desktop, printer, Kindle, iBook, etc.)
- Bitonal 300 dpi
- OCR for full-text searching
- Multiple items batched as single PDF
- Minimal metadata

Reformatted Materials

- Tied & returned to shelves
- Foldered & added to archival collections
- Recycled

Important Reminders

- Digitization is a recovery of information
- Preservation ≠ Permanent
- Important ≠ Unique
- Commitment to value, not format

No Grandfather Clause

- Reasonable, not conceivable, expectations
- The keeping of records is primarily an act of faith