

Mass digitization at the UM Library

Emerged projects around the
Google partnership

Library-wide

- Coordinated by, not contained within, IT
 - Greater interoperation between IT units
 - Touches on most everyone in the library
 - *We've tried to adapt the organization to the project, not vice versa.*
-
- Consequently...I'm speaking for a lot of people and am not necessarily authoritative. ;^)

Scaling storage

- RFP process concluded this summer
- Over a dozen suppliers; selection based on
 - Scalability of storage *and* performance
 - Ease of management
- Two Isilon NAS clusters, currently 100 TB in 32 RU (4.5 TB/RU), scalable to 1 PB

Scaling ingest

- Download/validation/ingest handled by GROOVE (code is available)
- Network-intensive: max 1 MB/s per download, need 3 MB/s across entire process
- Other parts CPU-intensive or I/O-intensive
- Added locking and multiprocessing to GROOVE
- Moved to dual dual core blade servers and NAS
- Running two instances; will scale up as needed once using Isilon storage

New features in MBooks

- Scanning brittle volumes (rejected by Google)
 - Refocused existing local scanning/vendor capacity
 - 15K volumes per year
 - Destructive scanning if not unique (>3 available)
 - Send to Google and route through existing processes
 - If not in print, make available online within library buildings (section 108)
- Services for visually impaired: list of students received from campus office; if student checks out book, also available online (section 107)
- Personalized/sharable/searchable collections

Copyright determination

- Tech services staff with special training use MBooks to check
 - Presence of printed copyright notice
 - Renewal records for 1923 – 1963
 - Government documents
- 1500 volumes cleared/month since June
- Interested in sharing and cross-checking results with other institutions

Shared Digital Repository

- Subsidized for-fee service offered by partnership of Indiana University and University of Michigan
- Vision: aggregated collection of digitized holdings from as many libraries as possible
- Seeded by CIC as founding charter members; contractual stage nearly finalized
- Will provide access as well as preservation, a la trusted digital repository principles
- Coordination with other services (GBS, etc)

- Big? Hopefully.

More information

- Find me to
 - Get access to GROOVE code
 - Hear how cool Isilon storage is
- I will point you to others for
 - More about MBooks and brittle books
 - How our copyright determination is done
 - Joining the SDR