Anatomy of a Digitization Project: Dissecting the Process

Mary E. Piorun
University of Massachusetts Medical School

Et al.
Anatomy of A Digitization Project: Dissecting the Process
Mary E. Piorun, MLIS, AHIP and Lisa A. Palmer, MLIS
The Lamar Soutter Library ~ University of Massachusetts Medical School

The Project: Digitizing 300 doctoral dissertations in-house and adding the full text to the institutional repository.

Method
The library purchased a license for ProQuest Digital Commons institutional repository software and named the school’s repository eScholarship@UMMS. A team was established with representatives from Systems, project management and technical support), Cataloging (metadata support), and Reference (outcome).

Evaluation
- Project under budget
- 325 dissertations submitted (most with full-text)
- Over 11,000 full text downloads in 11 months
- Increased visibility on campus
- Led to other successful projects
- International Medical Education Reports
- Senior Scholar projects

Conclusion
All available dissertations have been digitized to date. This project’s success is due to a number of factors: library funding, support, management, and team skills; buying a hosted product; and the support of the Graduate School Dean. Future success of the repository will depend on continued funding, dedicated repository staff, increased faculty and department participation, and greater campus awareness.

Future Directions & Challenges
Administrative: Copyright issues, workflow and time management, marketing and promotion.
Content Recruitment: Identifying new partners and processing articles from open access journals.

1: Getting Started
Before any work could be done a number of issues had to be resolved:

Coordination: Systems and Cataloging developed departmental workflows and jointly coordinated a process to handle materials and maintain communication.

Permissions: Alumni were contacted by e-mail or mail and asked to sign a form granting the library permission to scan their dissertation. A database was created to track progress.

Metadata: The team reviewed the template for adding new records to the system. New fields were added and existing fields were enhanced. For all views of the data, the display order of fields was specified. Catalogers wrote data entry instructions and determined the process for re-utilizing MARC metadata from the online catalog.

2: Scanning
Scanning is done by Circulation staff at night and on the weekends in the ILL work-area. Dissertations average about 250 pages in length and are single-sided. On average 2 dissertations per night are scanned and 5 on weekends.
A Canon Image Runner 3300 black & white scanner/copier/printer is used along with eCopy 3.1 software.

3: Quality Control
Quality control will re-assemble the scanned files into one and spot check for completeness. The black edges from scanning are removed up to chapter one and the quality of any images are reviewed. The file is then saved in various file formats. All of this work is done in the eCopy software.

4: Table of Contents
In order to enhance the quality of the final dissertation file we create a table of contents in the PDF. This is done using the bookmarks feature. We bookmark the following areas:
- Title
- Signature
- Abstract
- Chapters
- References

5: OCR Abstract
Using e-Copy’s OCR technology the abstract is saved as a text file. The file is reviewed for corrections and cleanup. Then HTML tagging is applied to display bold text, italics, superscripts, subscripts and any scientific characters.

6: Add to eScholarship@UMMS
The catalogers add records for the scanned dissertations with abstracts into eScholarship@UMMS. Some information – author, title, subjects – is copied and pasted from the library’s online catalog. Other information must be added directly. The catalogers upload the PDF of the full text and include a link to the record in the online catalog so that users can easily locate a paper version of the dissertation if desired.

The alumni database is updated with the date the dissertation was added into eScholarship@UMMS.

If an author has not yet granted permission, the dissertation still goes through the scanning, quality control, table of contents, and abstract processes. A record is added to eScholarship@UMMS, but the full text is not uploaded until permission is received.

http://escholarship.umassmed.edu