

7-1-2008

# Digitizing Dissertations for an Institutional Repository: A Process and Cost Analysis

Mary E. Piorun

*University of Massachusetts Medical School, mary.piorun@umassmed.edu*

Lisa A. Palmer

*University of Massachusetts Medical School, lisa.palmer@umassmed.edu*

Follow this and additional works at: [http://escholarship.umassmed.edu/lib\\_articles](http://escholarship.umassmed.edu/lib_articles)



Part of the [Library and Information Science Commons](#)

---

## Repository Citation

Piorun, Mary E. and Palmer, Lisa A., "Digitizing Dissertations for an Institutional Repository: A Process and Cost Analysis" (2008).  
University of Massachusetts Medical School. *Library Publications and Presentations*. Paper 94.  
[http://escholarship.umassmed.edu/lib\\_articles/94](http://escholarship.umassmed.edu/lib_articles/94)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Library Publications and Presentations by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

# Digitizing dissertations for an institutional repository: a process and cost analysis\*

Mary Piorun, MSLS, MBA, AHIP; Lisa A. Palmer, MSLS

See end of article for author's affiliations.

DOI: 10.3163/1536-5050.96.3.008

**Objective:** This paper describes the Lamar Soutter Library's process and costs associated with digitizing 300 doctoral dissertations for a newly implemented institutional repository at the University of Massachusetts Medical School.

**Methodology:** Project tasks included identifying metadata elements, obtaining and tracking permissions, converting the dissertations to an electronic format, and coordinating workflow between library departments. Each dissertation was scanned, reviewed for quality control, enhanced with a table of contents, processed through an optical character recognition function, and added to the institutional repository.

**Results:** Three hundred and twenty dissertations were digitized and added to the repository for a cost

of \$23,562, or \$0.28 per page. Seventy-four percent of the authors who were contacted (n=282) granted permission to digitize their dissertations. Processing time per title was 170 minutes, for a total processing time of 906 hours. In the first 17 months, full-text dissertations in the collection were downloaded 17,555 times.

**Conclusion:** Locally digitizing dissertations or other scholarly works for inclusion in institutional repositories can be cost effective, especially if small, defined projects are chosen. A successful project serves as an excellent recruitment strategy for the institutional repository and helps libraries build new relationships. Challenges include workflow, cost, policy development, and copyright permissions.

## INTRODUCTION

Digitization projects in libraries seem ubiquitous as libraries become increasingly involved in the acquisition, development, and management of digital information [1]. Libraries typically target archival and special collections materials such as historical documents and photographs [2]. Projects to digitize vast collections of books began as early as 1971 with Project Gutenberg and are now getting widespread media attention with the launch of Google Book Search, the Internet Archive, and others [3]. In an April 2007 list of ten assumptions about the future that would significantly impact academic libraries and librarians, the Association of College & Research Libraries Research Committee placed digitization at the top of the list, stating, "There will be an increased emphasis on digitizing collections, preserving digital archives, and improving methods of data storage and retrieval" [4].

A related emergent trend in academic libraries is the implementation of institutional repositories (IRs), digital collections that capture and preserve the intellectual output of university communities [5]. A search of *OpenDOAR*, the Directory of Open Access Repositories, lists 298 academic repositories in North America [6]. Health sciences libraries are among those contributing to this trend; of 125 libraries that responded to a 2006 supplementary survey for the Annual Statistics of Medical School Libraries in the United States and Canada, 28 have established IRs

### Highlights

- The Lamar Soutter Library partnered with the University of Massachusetts Medical School Graduate School of Biomedical Sciences to digitize doctoral dissertations for inclusion in a newly created institutional repository.
- Seventy-four percent of dissertation authors (209/282) gave permission for the digitization. The cost to process the entire dissertation collection in-house was \$23,562, only \$1,062 more than the estimate to outsource.
- Digitizing the dissertation collection increased access: the print collection was used 723 times in the past 5 years, while the electronic collection was used 17,555 times in 17 months.

### Implications

- Digitizing student works is an effective way to begin populating an institutional repository.
- In-house digitization projects can be cost-competitive with outsourced alternatives.
- A repository can be a catalyst for developing relationships in the institution by providing the library with a new avenue for outreach.
- Skills and experience gained from a small project can be applied to larger-scale projects.

\* Based on a poster at MLA '07, the 107th Annual Meeting of the Medical Library Association; Philadelphia, PA; May 20, 2007; and a presentation at the Scanning Forum 2006; Charlottesville, VA; November 6, 2006.

and 70 are planning to add or are considering offering a repository [7]. According to Foster and Gibbons, libraries build IRs because they “provide an institution with a mechanism to showcase its scholarly output, centralize and introduce efficiencies to the stewardship of digital documents of value, and respond proactively to the escalating crisis in scholarly communication” [8].

Medical librarians are just beginning to report their experiences with institutional repositories in the professional literature [9–13]. In one case study, Krevit and Crays [13] describe challenges that the Texas Medical Center experienced in piloting a multi-institutional repository, including copyright concerns and lack of faculty participation. An analysis by Singarella and Schoening [14] of the surveys conducted between 2005 and 2007 by the Association of Academic Health Sciences Libraries and a survey conducted in 2006 by the Association of Research Libraries [15] confirmed that the challenges experienced at the Texas Medical Center were not unique. Libraries are the drivers of IRs at their institutions, as few faculty members identify and self-archive their own materials. Libraries struggle to recruit content and employ a variety of strategies to enlist submissions [16–19]. Content may vary, but a recent study by McDowell reports that student works account for the largest percentage of documents in institutional repositories, approximately 41.5% [19].

The following case study describes a nexus of these two trends: digitization of student scholarly works and institutional repositories. The first digitization project for the Lamar Soutter Library at the University of Massachusetts Medical School (UMMS) was to digitize 300 doctoral dissertations and add the full text to the school’s new IR.

## BACKGROUND

Founded in 1970, UMMS encompasses the graduate schools of medicine, nursing, and biomedical sciences. The Lamar Soutter Library holds 175,000 print volumes and provides access to 316 databases, 4,650 electronic journals, and 359 electronic books. The IR is the library’s first comprehensive digital initiative.

In early 2006, the library purchased a license for ProQuest Digital Commons,† a hosted institutional repository system, and named the repository “eScholarship@UMMS” <<http://escholarship.umassmed.edu>>. The team implementing the repository, a previously reported process [12], consisted of representatives from the library’s systems (project management and technical support), cataloging (metadata support), and reference (outreach) departments. In March 2006, the dean of the UMMS Graduate School of Biomedical Sciences (GSBS) expressed interest in

digitizing the school’s dissertations. The GSBS had produced 300 dissertations, most of which were available only in print format. The team thought this would be an excellent demonstration project: it was supported by the dean, it was a manageable size, metadata could be reutilized from the library’s online public access catalog (OPAC), and the dissertation authors held the copyright. In May 2006, the library and GSBS partnered to make the dissertations fully searchable on the web.

## METHODOLOGY

### Outsourcing versus insourcing

The team investigated 2 options for digitizing the dissertations: outsourcing to UMI or performing the work in-house. UMI estimated the cost to be \$75 per title (\$22,500 total) and 8–12 weeks processing time. The basis for the library estimate was created by library staff scanning and locally preparing 3 sample dissertations. Table 1 shows the library’s cost estimate of \$27,750—for staffing, project management, equipment, and software—and 725 hours of processing time (or 18 weeks when represented as a 40-hour work week). In all instances, except for project management, the team assumed the work would be performed by temporary help. The team had 2 issues of concern: at the time, electronic files created by UMI were not full-text searchable, and the graduate school would need to commit to sending all future dissertations to UMI to keep the database current.

The project team recommended that the library process the dissertations in-house, despite longer time to process and higher cost, in order to gain experience, retain access to materials throughout the project, and have tighter control over scanning quality. Library administration accepted the recommendation to do the digitization locally, citing “gaining experience” as the major benefit; however, \$27,750 was not available to fund the project. Ten thousand dollars was allotted to hire temporary staff, with the understanding that circulation staff and interlibrary loan equipment would be utilized for scanning and team catalogers would add dissertations to the repository. It was also recognized that the project could not be completed in 18 weeks as staff assigned to the project would need to incorporate the dissertation tasks into their daily workload.

### Metadata

To fully utilize metadata from the library’s integrated library system, team catalogers customized default templates in the Digital Commons software designed to control the indexing and display of a collection of records. Customizations were necessary to fully describe the dissertations and incorporated features such as the activation of live link functionality in fields where uniform resource locators (URLs) might be included, the addition of a field to record authors’ UMMS departmental affiliations, and the accommo-

† In July 2007, Berkeley Electronic Press (bepress), the original developers of the software, resumed full support of the Digital Commons product. It is now called bepress Institutional Repository. For more information, see the product description available at <<http://www.bepress.com/ir/>>.

**Table 1**  
Estimate versus actual costs and processing times

Tasks	Estimate (300 documents)						Actual (320 documents)					
	Hours	Rate (\$)	Labor cost/ cost (\$)	Min per doc	Cost per doc (\$)	Cost per page (\$)*	Hours	Rate (\$)	Labor cost (\$)	Min per doc	Cost per doc (\$)	Cost per page (\$)*
Scanning	225	20	4,500	45	15.00	0.06	240	20	4,800	45	15.00	0.06
Quality control	225	25	5,625	45	18.75	0.08	133	22	2,926	25	9.14	0.04
Optical character recognition (OCR) of the abstract	100	25	2,500	20	8.33	0.03	160	22	3,520	30	11.00	0.04
Add to institutional repository (IR)	100	25	2,500	20	8.33	0.03	54	35	1,890	10	5.91	0.02
Signature page						0.00	133	22	2,926	25	9.14	0.04
Replace file						0.00	26	35	910	5	2.84	0.01
Project management	75	35	2,625	15	8.75	0.04	160	35	5,600	30	17.50	0.07
Equipment/software			10,000						990			
Total	725		27,750	145	59.17	0.24	906		23,562	170	70.54	0.28

\* Page length is estimated at 250 per document.

dation of Medical Subject Headings by changing the field delimiter from a comma to a semicolon. For instance, "Libraries, Medical" and "Library Technical Services" previously displayed as "Libraries" and "Medical; Library Technical Services." Catalogers copied and pasted title and subject data from the OPAC into the repository manually, using macros when possible. Though the Digital Commons software contained a batch loader functionality, it was not used in the submission process due to the batch loader having a separate extensible markup language (XML) schema that at the time could not be programmed to match the customized dissertation templates.

### Digitization and submission process

Using alumni contact data provided by the graduate school, library staff wrote to the dissertation authors to request copyright and digitization permissions. Alumni were asked to grant permission immediately, while current graduates were given the option to add only an abstract and delay adding the full-text for one year to allow for publishing opportunities. Initially, only dissertations for which the library secured permissions were scanned and processed. Once those were completed, a decision was made by the project team to scan the remaining dissertations, add records with the abstracts to the repository, and store the full-text files until permission was obtained. Dissertations averaged 250 pages in length and were single-sided, with a mix of text, tables, graphs, and images. They were scanned using a Canon Image Runner 3,300 with eCopy version 3.1, a software program used for scanning, optical character recognition (OCR), and portable document format (PDF) creation. Figures 1 and 2 illustrate the digitization process. Figure 3 shows a typical dissertation record in eScholarship@UMMS.

### An unexpected step to alleviate privacy concerns

As the project neared completion, the dean of the graduate school expressed concern about the signature pages of the dissertations being made public. The team asked ProQuest's UMI Dissertation Publishing

its policy on this issue and learned UMI stopped scanning signature pages in 2005. The team concluded it was worth the additional time and cost to re-create a "blank" signature page for each dissertation, which would retain the names of the advisor and review committee without their signatures (this information was not stored elsewhere). The new signature pages were created and reinserted into the PDF files. Cataloging staff then substituted the revised PDFs in eScholarship@UMMS.

### RESULTS

The total number of documents processed was 320, 300 from previous graduates and 20 dissertations submitted by an additional 20 students over the course of the project. The project team was able to successfully contact 282 of the 320 authors, and 209 (74%, 209/282) granted permission to digitize their dissertations. The dissertations (or records providing abstracts only in cases where permission was not granted) were all available online by March 2007.

### Processing time

Actual processing times are summarized in Table 1. The total hours to process the materials were 906 hours, exceeding the original estimate of 725 hours by 181 hours. One-hundred and fifty-nine hours of this difference can be attributed to the unexpected need to replace the signature pages in each dissertation. The total duration of the project was 12 months, as the circulation staff members who scanned the dissertations were not assigned to the project full time. They scanned on average 2 dissertations per night and 5 on weekends. Spreading the work over the course of 1 year allowed for multiple attempts to contact alumni for permission.

Closer analysis of the estimated and actual time needed per dissertation shows 2 important factors. First, the initial time estimate to process a dissertation was low (145 minutes vs. 170 minutes); however, if the additional step of replacing the signature pages was not required, the original estimate would have been accurate. Second, regardless of the difference in the total time needed per dissertation, some important

**Figure 1**  
Digitization process

- **Quality control:** Page orientation, page order, completeness, and image quality were reviewed and corrected. A searchable PDF was created from the eCopy file to be used throughout the rest of the process.
- **Table of contents:** Each major section of the dissertation received an entry in the table of contents, including title page, list of figures, and chapters.
- **Optical character recognition (OCR) abstract:** An image-based version of the abstract was created using eCopy and processed through the OCR function, then exported as a text file. The text file was then examined for capitalization, punctuation, spacing, and scientific notations; marked up in HTML to preserve original layout; and used by the cataloger adding the metadata for the dissertation.
- **Addition to eScholarship:** Catalogers reutilized data from the library's online public access catalog to create the repository record. Information pertaining to copyright, quality of the scanned images, and missing pages was added as applicable. The HTML version of the abstract was copied into the record, as well as a link to the library's holdings information for the dissertation. The full-text PDF was subsequently added if the author granted permission. For dissertations with an abstract only, the cataloger added a comment to the record stating the library was in the process of obtaining permission from the author. When permission was obtained, the record in eScholarship@UMMS was revised to add the full-text PDF and edit the Comments field. The catalogers also added a link from the library catalog to the repository.

areas were underestimated, such as the time to OCR the abstract and overall project management. Issues that contributed to this miscalculation include the extra time to correct the scientific notation in the OCR process and the total project management time required to obtain permissions from authors to digitize their work.

**Equipment and software**

The work was accomplished using existing library scanning equipment. The library already owned copies of the software used throughout the process: Microsoft Access, eCopy, Adobe Acrobat, and Adobe Illustrator. Because eCopy came with a scaled down version of the Readiris OCR software, the library purchased 3 copies of the full Readiris program for a total of \$990; however, these were not used in the project because the 2 versions conflicted. Thus, the original estimate of \$10,000 for equipment and software was too high.

**Labor**

Actual labor costs, as shown in Table 1, were \$22,572 versus the estimated costs of \$17,750. In the initial estimate, a temporary worker was assigned the task of adding the dissertation to the repository; however, 2

staff catalogers performed this work at a resulting higher rate. The \$10,000 allotted for a temporary worker paid for quality control, OCR work, and editing of the signature pages for a total cost of \$9,372. This labor cost would have been \$6,446 if the extra step of editing the signature pages had not been necessary.

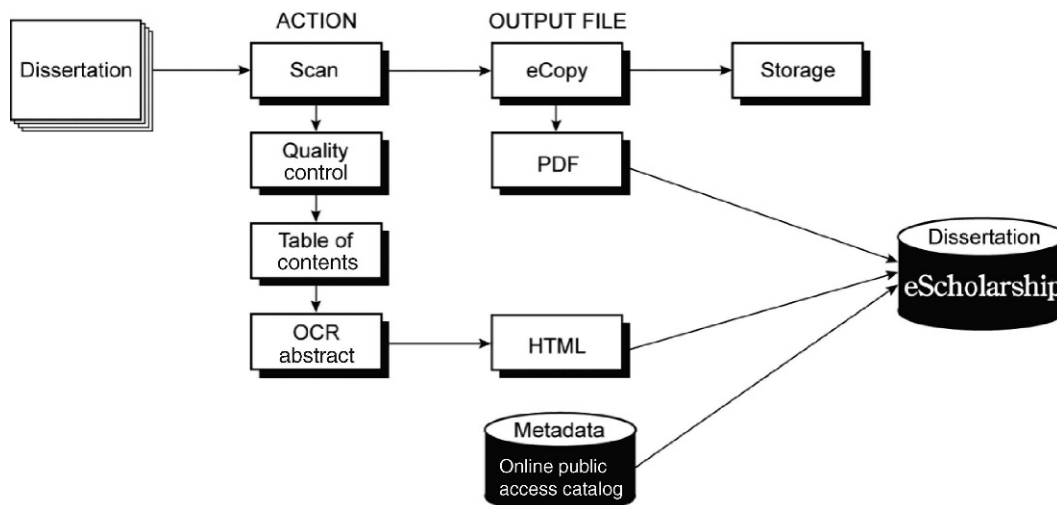
**Budget**

Total project costs were \$23,562 (\$990 software, \$22,572 labor) or \$0.28 per page (Table 1). This is \$4,188 less than the original estimate of \$27,750 to process the dissertations in-house and \$1,062 more than the estimated cost to outsource the dissertations to UMI.

**Usage**

Historical circulation data from May 1999 through November 2007 show the library's print dissertation collection was used 723 times. This is in stark contrast to the first 17 months the electronic collection was available (June 2006 through November 2007). Downloads of full-text PDF dissertations from eScholarship@UMMS totaled 17,555, with 10,497 originating from Google searches.

**Figure 2**  
Process to convert and add dissertations to repository



**Figure 3**  
Typical dissertation record

---

**Graduate School of Biomedical Sciences**

---

GSBS Dissertations [\[Browse Contents\]](#) [\[Search\]](#) [\[gsbs Website\]](#) [\[Submit a Paper\]](#)

<Previous Dissertation, Doctoral Next Dissertation, Doctoral>

**TITLE:**  
Functions of the Cdc14-Family Phosphatase Clp1p in the Cell Cycle Regulation of *Schizosaccharomyces pombe*: A Dissertation

**AUTHOR(S):**  
Susanne Trautmann, [University of Massachusetts Medical School](#)

**DATE:** 05/20/05

**UMMS AFFILIATION:** Graduate School of Biomedical Sciences, Molecular Genetics & Microbiology, Interdisciplinary Graduate Program

**DOCUMENT TYPE:** Dissertation, Doctoral

**SUBJECTS:** Cytokinesis; Cell Cycle Proteins; Gene Expression Regulation, Enzymologic; Protein-Serine-Threonine Kinases; Schizosaccharomyces pombe Proteins; Genes, cdc; Academic Dissertations; Dissertations, UMMS

■ [Download the Document](#) (PDF format - 7.5 MB) - May 2005

■ **Related Files:** trautmann video1.mov (569 kB)  
*Video1: S. pombe cells expressing clp1-GFP sid4-GFP*  
 trautmann video2.mov (991 kB)  
*Video2: S. pombe cells expressing sid4-GFP*  
 trautmann video3.mov (546 kB)  
*Video3: S. pombe cells with GFP labeled centromere II (cen2-GFP), released from nda3-km311 block*  
 trautmann video4.mov (1018 kB)  
*Video4: S. pombe cells with GFP labeled centromere II (cen2-GFP) and deletion of clp1, released from nda3-km311 block*

■ Tell a colleague about it.

**ABSTRACT:**  
 In order to generate healthy daughter cells, nuclear division and cytokinesis need to be coordinated. Premature division of the cytoplasm in the absence of chromosome segregation or nuclear proliferation without cytokinesis might lead to aneuploidy and cancer. Cdc14-family phosphatases are highly conserved from yeast to humans, but were only characterized in *Saccharomyces cerevisiae* at the time this thesis was initiated. Cdc14 had been identified as the effector of a signaling cascade homologous to the SIN, called the mitotic exit network (MEN), which is required for exit from mitosis. This thesis describes the identification of the *S. pombe* Cdc14-like phosphatase Clp1p as a component of the cytokinesis checkpoint. Clp1p opposes CDK activity, and Clp1p and the SIN activate each other in a positive feedback loop. This maintains an active cytokinesis checkpoint and delays mitotic entry. We further found that Clp1p regulates chromosome segregation.  
 Concluding, this thesis describes discoveries adding to the characterization of the cytokinesis checkpoint and the function of Clp1p. While others found that Cdc14-family phosphatases, including Clp1p, have similar catalytic functions, we show that their biological function may be quite different between organisms, possibly due to different biological challenges.

**COMMENTS:**  
 Chapter 5 not included in digitized version, per author's request.

**RELATED RESOURCES:** [Link to record for print version in Library Catalog](#)

---

This abstract has been cut short for display purposes. This particular dissertation was accompanied by a disc containing four videos that were linked as supplementary files. The Comments field notes that one chapter was not digitized at the author's request, because she was in the process of publishing an article based on that chapter. The Related Resources field links to the OPAC record describing the print version of this dissertation and displaying its availability.

---

## Staff development

Team members became more familiar with the repository software, metadata standards, scanning, and OCR technologies and developed closer working relationships on the team and between departments. The team developed a greater awareness of the importance of copyright compliance.

## CONCLUSION

Many libraries have viewed digitizing collections as too expensive an undertaking in this time of tight budgets [20]. Chapmen of Harvard University states the costs for scanning, OCR, and quality control work can be as much as 48% of a project's total costs [21]. Equivalent costs throughout the Lamar Soutter Library's dissertation project match this estimate (47.79%). Using Chapmen's group of activities—scanning, OCR, and quality control—the per-page cost to process black-and-white text in a bound volume can range from \$0.10 to \$1.40 [22, 23]. Both these figures are based on outsourcing the work. The Lamar Soutter Library's internal costs were competitive with these estimates, at \$0.28 per page. This suggests the cost to digitize may be within the reach of many medical libraries and a viable option to populate institutional repositories.

The usage statistics for this collection indicate that by disseminating the dissertations through eScholarship@UMMS, which is indexed by Google, access and use increased substantially. Studies indicate that individuals who publish their research online in addition to publishing in traditional scholarly venues are cited more often than those who rely solely on paper publications [24–27]. In digitizing the GSBS dissertations, the library has assisted in making the school's research more widely available.

The team faced challenges such as workflow, cost concerns, policy development, and permissions. Communication and coordination between internal and external departments was vital and minimized errors. As the team learned, regardless of the amount of planning and thought that goes into a project, there is always the possibility that each record or file will need to be reworked. Decisions made in processing the dissertations set a precedent for future collections, such as adding documents without the full text if permission has not been obtained. The team acknowledges this could result in user frustration because they cannot get access to the full text. The team has worked hard to contact as many dissertation authors as possible to keep incomplete records to a minimum.

Nolen and Costanza described their experience in populating the repository at Trinity University, which

also focused on student works, by noting, "it's important to start small, choosing projects that have usefulness to our constituents" [28]. The Lamar Soutter Library also found having a small, defined project had many benefits. It allowed the team to experience an early success and manage staff and resources by gradually incorporating the work. The team also gained experience with Digital Commons, metadata standards, and copyright. Additionally, this project served as a recruitment strategy to other campus departments through coordinated promotion by GSBS and the library for further population of the institutional repository. New materials recruited include student works, nursing dissertations, and faculty publications, a small portion of which required digitization.

For UMMS, digitizing dissertations proved to be a successful and cost-effective recruitment strategy and helped the library build stronger relationships at the medical school to secure future content. The team's quick response to the dean's privacy concerns built a foundation of trust for future work. Currently, all dissertations are submitted to the library in both print and electronic format along with a signed permission form to digitize the work. The library anticipates that building this relationship with students will make it easier to recruit future scholarly works over the life of a researcher's career at the medical school.

## REFERENCES

- Muir A. Preservation, access and intellectual property rights challenges for libraries in the digital environment [Internet]. London, UK: Institute for Public Policy Research; 2006 [rev. 5 Jun 2006; cited 21 Nov 2007]. <<http://www.ippr.org/publicationsandreports/publication.asp?id=464>>.
- Institute of Museum and Library Services. Status of technology and digitization in the nation's museums and libraries [Internet]. Washington, DC: The Institute; 2002 [rev. 23 May 2002; cited 21 Nov 2007]. <<http://www.ims.gov/resources/TechDig02/2002Report.pdf>>.
- Coyle K. Mass digitization of books. *J Acad Librariansh*. 2006;32(6):641-5.
- Mullins JL, Allen FR, Hufford JR. Top ten assumptions for the future of academic libraries and librarians: a report from the ACRL Research Committee. *C&RL News*. 2007 Apr;68(4), (Available from: <<http://www.lita.org/ala/acrl/acrlpubs/crlnews/backissues2007/april07/tenassumptions.cfm>>. [cited 21 Nov 2007].)
- Crow R. The case for institutional repositories: a SPARC position paper. release 1.0 [Internet]. Washington, DC: Scholarly Publishing and Academic Resources Coalition; 2002 [rev. Aug 2002; cited 1 Aug 2007]. <[http://www.arl.org/sparc/bm~doc/ir\\_final\\_release\\_102.pdf](http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf)>.
- University of Nottingham. OpenDOAR: the directory of open access repositories [Internet]. Nottingham, UK: The University; 2007 [rev. 26 Nov 2007; cited 26 Nov 2007]. <<http://www.opendoar.org>>.
- Association of Academic Health Sciences Libraries. Annual statistics of medical school libraries in the United States and Canada. [Internet]. The Association [password protected; cited 3 Aug 2007]. <<http://aahsl.ccr.buffalo.edu>>.
- Foster NF, Gibbon S. Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Mag* [Internet]. 2005 Jan;11(1) [cited 5 Oct 2007]. <<http://www.dlib.org/dlib/january05/foster/01foster.html>>.
- Phillips H, Carr R, Teal J. Leading roles for reference librarians in institutional repositories: one library's experience. *Ref Serv Rev*. 2005;33(3):301-11.
- Hatfield AJ, Kelley SD. Case study: lessons learned through digitizing the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research Collection. *J Med Libr Assoc*. 2007 Jul;95(3):267-70.
- Mower A. Developing an institutional repository: an insider's look at the University of Utah IR. *Libr Student J* [Internet]. 2006 Nov [cited 20 Sep 2007]. <<http://informatics.buffalo.edu/org/ljsj/>>.
- Piorun ME, Palmer LA, Comes JF. Challenges and lessons learned: moving from image database to institutional repository. *OCLC Syst Serv*. 2007;23(2):148-57. DOI: 10.1108/10650750710748450.
- Kreivit L, Crays L. Herding cats: designing Digital Commons @ The Texas Medical Center, a multi-institutional repository. *OCLC Syst Serv*. 2007;23(2):116-24. DOI: 10.1108/1065075071074844.
- Association of Academic Health Sciences Libraries. Annual statistics of medical school libraries in the United States and Canada. [Internet]. The Association. Singarella T, Schoening P. AAHSL institutional repositories (IR) survey summary and analysis: 2005-2007 comparison. [password protected; cited 18 Sep 2007]. <<http://aahsl.ccr.buffalo.edu>>.
- Bailey C. Institutional repositories. Washington, DC: Association of Research Libraries; 2006.
- Lynch CA, Lippincott JK. Institutional deployment in the United States as of early 2005 [Internet]. *D-Lib Mag* [Internet]. 2005 Sep;11(9) [cited 1 Aug 2007]. <<http://www.dlib.org/dlib/september05/lynch/09lynch.html>>.
- Markey K, Rieh SY, St. Jean B, Kim J, Yakel E. Census of institutional repositories in the United States: MIRACLE project research findings [Internet]. Washington, DC: Council on Library and Information Resources; 2007. [rev. Feb 2007; cited 19 Oct 2007]. <<http://www.clir.org/pubs/reports/pub140/pub140.pdf>>.
- Davis PM, Connolly M. Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Mag* [Internet]. 2007 Mar/Apr;13(3/4) [cited 5 Sep 2007]. <<http://www.dlib.org/dlib/march07/davis/03davis.html>>.
- McDowell CS. Evaluating institutional repository development in American academe since early 2005: repositories by the numbers, part 2. *D-Lib Mag* [Internet]. 2007 Sep/Oct;13(9/10) [cited 10 Oct 2007]. <<http://www.dlib.org/dlib/september07/mcdowell/09mcdowell.html>>.
- Institute of Museum and Library Services. Status of technology and digitization in the nation's museums and libraries [Internet]. Washington, DC: The Institute; 2006 [rev. Jan 2006; cited 21 Nov 2007]. <<http://www.ims.gov/resources/TechDig05/Technology%2BDigitization.pdf>>.
- Chapman S. Managing text digitization: making good digital text objects. Presented at: School for Scanning: Building Good Digital Collections; Jun 2, 2005; Boston, MA.
- Boston Library Consortium. News release: Boston Library Consortium partners with Open Content Alliance to provide public access to digitized books [Internet]. The Consortium. [rev. 25 Sep 2007; cited 1 Oct 2007]. <[http://www.blc.org/news/blc\\_oca\\_release.html](http://www.blc.org/news/blc_oca_release.html)>.
- Kenny AR, Rieger OY. Moving theory into practice: digital imaging for libraries and archives. Mountain View, CA: Research Libraries Group; 2000.

24. Lawrence S. Free online availability substantially increases a paper's impact. *Nature*. 2001 May 31;411(6837):521.
25. Antelman K. Do open access articles have a greater research impact? *Coll Res Libr*. 2004;65(1):372–82. (Available from: <<http://www.ala.org/ala/acrl/acrlpubs/crljournal/crl2004/crlseptember/antelman.pdf>>. [cited 12 Oct 2007].)
26. Eysenbach G. Citation advantage of open access articles. *PLOS Biol*. 2006 May;4(5) [cited 26 Nov 2007]. DOI: 10.1371/journal.pbio.0040157.
27. Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PLOS One*. 2007 Mar;2(3) [cited 26 Nov 2007]. DOI: 10.1371/journal.pone.0000308.
28. Nolan C, Costanza J. Promoting and archiving student work through an institutional repository: Trinity University, LASR, and Digital Commons [Internet]. San Antonio, TX:

Trinity University; 2006 [rev. Jun 2006; cited 18 Sep 2007]. <<http://hdl.handle.net/10090/502>>.

#### AUTHORS' AFFILIATIONS

**Mary Piorun, MSLS, MBA, AHIP**, [mary.piorun@umassmed.edu](mailto:mary.piorun@umassmed.edu); Associate Director for Technology Initiatives and Resource Management; **Lisa A. Palmer, MSLS**, [lisa.palmer@umassmed.edu](mailto:lisa.palmer@umassmed.edu), Catalog Librarian; Lamar Soutter Library, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655

*Received October 2007; accepted December 2007*