

2008-10-24

# Digitizing Dissertations for the eScholarship@UMMS Institutional Repository

Mary E. Piorun

*University of Massachusetts Medical School*

Lisa A. Palmer

*University of Massachusetts Medical School*

Follow this and additional works at: [https://escholarship.umassmed.edu/lib\\_articles](https://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 the eScholarship@UMMS Institutional Repository" (2008).  
University of Massachusetts Medical School. *Library Publications and Presentations*. Paper 96.  
[https://escholarship.umassmed.edu/lib\\_articles/96](https://escholarship.umassmed.edu/lib_articles/96)

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 the eScholarship@UMMS Institutional Repository

Lisa Palmer & Mary Piorun

Lamar Soutter Library  
University of Massachusetts Medical School  
Worcester, MA

<http://library.umassmed.edu>



# Overview

---

- Background
  - 1<sup>st</sup> digitization project
  - Team members and roles
  - Choosing a repository system
  - Identifying manageable first project
- Project: digitizing 300 dissertations in-house 2006-2007
  - Partnership with one of our graduate schools
  - Metadata
  - Permissions process
  - Technical decisions
  - Workflow
  - Skills needed
  - Coordination between and within departments
- Ongoing operations

# UMass Medical School



- Massachusetts' only public medical school, founded in 1970
- Consistently ranked in top 10 in primary care education among 125 U.S. medical schools by U.S. News & World Report
- 990 students and growing
- School of Medicine, Graduate School of Nursing, Graduate School of Biomedical Sciences
- Clinical partner: UMass Memorial Health Care
- Separate graduate campus in UMass system

# Lamar Soutter Library



- National Library of Medicine's Regional Medical Library for the New England Region
- 210,000 volumes
- Journals: 1350 print subscriptions, 5000 subject-focused electronic subscriptions
- Special collections: rare books, history of medicine, humanities in medicine, consumer health, pediatrics, archives
- 42 Library FTE
- Medium-sized academic health sciences library

# Team Charge

---

“Investigate institutional repository products and make a recommendation for the Medical School”

# The Team

---

- Associate Director for Technology Initiatives and Resource Management (Project Leader): Project management, technology, workflow analysis, usability
- Associate Director for Education and Research Services: Outreach to faculty and students, copyright, training
- Catalog Librarians (2): Metadata, indexing, documentation, quality control, usability

# System Evaluation

---

- Research: articles, discussion lists, library websites, users from other libraries, workshops, product demonstrations
- “Score card”



# The Score Card

<b>User Interface: 25 points</b> <ul style="list-style-type: none"><li>▪ Customizability</li><li>▪ User friendliness</li><li>▪ Searching/retrieval</li><li>▪ Submission process</li><li>▪ Navigation</li></ul>	<b>Cost: 10 points</b> <ul style="list-style-type: none"><li>▪ Initial cost</li><li>▪ Annual maintenance fee</li><li>▪ Licensing fee</li><li>▪ Impact on staffing models</li><li>▪ Pricing model</li></ul>
<b>Tools: 30 points</b> <ul style="list-style-type: none"><li>▪ E-mail lists</li><li>▪ Faculty web pages</li><li>▪ E-journal publication</li><li>▪ Alerting service</li><li>▪ Controlled vocabulary lists</li><li>▪ Data feeds</li><li>▪ PDF conversion</li><li>▪ Ability to link related files</li></ul>	<b>Administration: 25 points</b> <ul style="list-style-type: none"><li>▪ Setup time</li><li>▪ Statistical reporting</li><li>▪ Interoperability/compatibility</li><li>▪ Maintenance interface</li><li>▪ Long-term maintenance required</li><li>▪ Accepted file formats</li><li>▪ Export of data</li><li>▪ New staff skills required</li><li>▪ Branding/customizing</li><li>▪ Training</li><li>▪ Access control</li></ul>
<b>Company/Community: 10 points</b> <ul style="list-style-type: none"><li>▪ Customer service/support</li><li>▪ User documentation</li><li>▪ Company stability</li><li>▪ Customer references</li><li>▪ Number of product installations</li><li>▪ Installed base</li></ul>	

# Digital Commons

<http://www.bepress.com/ir/>

---

- 2-year license purchased in January 2006
- Hosted
- Cool stuff: ability to link video & sound files, data sets
- OAI compliant
- Usage statistics, including monthly readership statistics emailed to authors
- Functionality that would make it easier to promote the repository: email alerts, “paper of the day”
- Faculty researcher pages, online journal publishing

# Getting Started

---

- “eScholarship@UMMS”
  - <http://escholarship.umassmed.edu>
- Testing with Library staff publications: articles, presentations & posters
- Basic customizations to end-user and administrative interfaces

# Pilot Project

---

Needed a manageable first project to

- Populate repository quickly
- Generate visibility
- Gain support

# Dissertations

---

- Graduate School of Biomedical Sciences, founded in 1979
- Good demonstration project
  - GSBS Dean interested in project
  - Reasonable number (~300)
  - Already cataloged and had metadata
- Very few submitted in electronic form
- Not submitted to UMI

# Preparing the System: Metadata

---

- “Supports export to XML Dublin Core (DC) format”
  - **not exactly**
    - Elements captured: title, creator, description, date, type, format, identifier, publisher, subject
    - Elements not captured: contributor, source, language, relation, coverage, rights
- Added new fields to dissertation template
  - DC elements not captured
  - Department (“UMMS Affiliation”)
  - ID number for bibliographic record in library catalog

# Sample XML Dublin Core View

```
<?xml version="1.0" encoding="utf-8" ?>
<dc>
<dc-record>
<title>Analysis of RNA Interference in <em>C. elegans</em>: A Dissertation</title>
<creator>Grishok, Alla</creator>
<description>RNA interference (RNAi) in the nematode Caenorhabditis elegans is a type of homology-dependent post-transcriptional gene silencing induced by dsRNA. This dissertation describes the genetic analysis of the RNA interference pathway and inheritance properties associated with this phenomenon. We demonstrate that the RNAi effect can be observed in the progeny of the injected animal for at least two generations. Transmission of the interference effect occurs through a dominant extragenic agent....</description>
<date>2001-09-27</date>
<type>text</type>
<format>application/pdf</format>
<identifier>http://escholarship.umassmed.edu/gpbs\_diss/139</identifier>
<publisher>eScholarship@UMMS</publisher>
<subject>Caenorhabditis elegans</subject>
<subject>RNA Interference</subject>
<subject>RNA, Small Interfering</subject>
<subject>Academic Dissertations</subject>
<subject>Dissertations, UMMS</subject>
</dc-record>
</dc>
```

# Preparing the System: Metadata (cont.)

---

- Added new document type:
  - Dissertation, Doctoral
- Activated live link functionality in Relation, Source, Comments fields
- Changed delimiter for subject field to accommodate MeSH

**Before:** <subject>Libraries</subject>

<subject>Medical; Library Technical Services</subject>

**After:** <subject>Libraries, Medical</subject>

<subject>Library Technical Services</subject>



# Preparing the System: Metadata (cont.)

---

- Specified display order of fields for various displays
- Data entry decisions
  - Use Relation field to provide link to record in OPAC for print version of dissertation
  - Use Rights field for information about copyright or permissions
  - Comments field

# Preparing the System: Metadata (cont.)

---

- How to re-utilize MARC data from online catalog
  - Small collection
  - Permission granted unevenly
  - Dismissed batch loader functionality
  - Decision: Copy & paste from OPAC; use macros where possible

# Sample Record

**Functions of the Cdc14-Family Phosphatase Clp1p in the Cell Cycle Regulation of *Schizosaccharomyces pombe***  
by **Susanne Trautmann**

**Comments** →

**Relation** →

**Video Files** →

Enter search terms:

in this series

[Advanced Search](#)

Notify me via email or RSS

#### Links

- Library
- Related Sites
- Help
- GSBS Dissertations Website

#### Browse

- Collections
- Disciplines
- Authors

#### Author Corner

- Author FAQ
- Submit Research

[Home](#) > [gsbs](#) > [GSBS DISS > 10](#) < [Previous Dissertation, Doctoral](#) [Next Dissertation, Doctoral](#) >

## GSBS DISSERTATIONS

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

[Download](#) 

[Tell a Colleague](#) 

[Print](#) 

[Download Adobe Reader](#) 

[BOOKMARK](#) 

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

#### Date

May 2005

#### 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

#### 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.

#### Comments

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

#### Related Resources

[Link to record for print version in Library Catalog](#)

#### Additional 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*

# Metadata Skills

---

Description  
Indexing  
Authority control  
Search and retrieval  
Testing  
Usability  
Quality control  
Documentation

# Outsource?

---

- UMI digitization service - \$22,500 (in 2006)
  - 2-3 month turn-around
  - Not full-text searchable
- In-house estimate
  - \$27,750 (in 2006)
    - Two temporary employees
    - Equipment
    - Project management
    - 14-week turnaround

# Estimates Per Title

	<b>Estimate (Minutes)</b>
<b>Scanning</b>	45
<b>Quality Control</b>	45
<b>OCR Abstract</b>	20
<b>Add to IR</b>	20
<b>Project Management</b>	15
<b>Total</b>	<b>145</b>

# Recommendation

---

- Process in-house
  - Gain experience
  - Retain access throughout
  - Tighter control
    - Project
    - Quality

# Given

---

- \$10,000 for temporary help
- Circulation staff
- ILL copier/scanners



# Process

---

1. Obtain Permission
2. Scan Dissertation
3. Quality Control
4. Build a Table of Contents
5. Process Abstract
6. Add Dissertation to eScholarship

# Permissions

---

- No process in place
  - Created two forms
    - Alumni
    - Current graduates
  - Forms approved by Legal department
- Contact 300 alumni
  - Access database
  - Local e-mail address

# Permissions Cont ...

---

- 310 authors
  - 250 contacted
- 167 granted permission
- 67% success rate

Skills: Access, Word, Mail Merge, Writing, Searching, Political

**Permissions** – Scanning – Quality – ToC – Abstract – Adding

# Scanning

---

- Who: Circulation staff
  - What: 250 pages – single-sided
  - When: Nights & weekends
  - Where: ILL office
- 
- Average: 2 per night, 5 on weekends

# Scanning

---

## ■ Hardware

- Canon networked printer/copier/scanner
- Image Runner 3300 black & white
- Image Runner C3200 color

## ■ Software

- ECopy 3.1
- <http://www.ecopy.com>

# Scanning

---

- Printout from OPAC
- Scan using eCopy
  - Break file up if called away
  - File is stored on the copier

Skills: Teamwork, Work Prioritization, Attention to Detail, Scanner Operation

Permissions – **Scanning** – Quality – ToC – Abstract – Adding

# Quality Control

---

- Assemble the files
- Check for completeness
- Clean up edges
- Verify image quality
- Saving of file in various formats

Skills: Attention to Detail, eCopy, Scanner, Save As, File Management

# Table of Contents

---

- In PDF using bookmarks to build a ToC
  - Title
  - Signature
  - Abstract
  - Chapters
  - References

Skills: Adobe Acrobat Professional, Bookmarks

Permissions – Scanning – Quality – ToC – Abstract – Adding



# Process Abstract

---

- eCopy
- OCR
- Notepad
- Cleanup
- HTML tagging
- Cataloging “In Box”

Skills: Attention to Detail, Proofreading, Basic HTML

Permissions – Scanning – Quality – ToC – **Abstract** – Adding

# Review: 3 Files

---

- eCopy file for future use
- Searchable PDF
- HTML version of abstract

# Add to eScholarship

---



Handoff to Cataloging

# Add to eScholarship

---

- Step 1: Add record to eScholarship
  - Copy/paste from OPAC
  - Author, title, department, date, subjects
  - Document type, comments, abstract, link to OPAC, upload PDF
- Step 2: Add full-text link to record in online catalog
- Step 3: Move files to “Added to eScholarship” folder
- Step 4: Update alumni database

Skills: Cataloging, Organization, Multi-tasking, HTML, Access, Teamwork

Permissions – Scanning – Quality – ToC – Abstract – **Adding**

# Decision

---

- No permission form on file
  - Scan dissertation and add record to eScholarship without full-text
- Pro: Process was working well, under budget
- Con: Adding records without the full text

# Workflow without Permission

---

- Step 1: Add record to eScholarship
  - Copy/paste from OPAC
  - Author, title, department, date, subjects
  - Document type, comments, abstract, link to OPAC
- Step 2: Add comment: seeking permissions
- Step 3: Move abstract to “Added to eScholarship” folder
- Step 4: Update alumni database

# When Permission is Obtained

---

- Step 1: In eScholarship, edit comments and upload PDF
- Step 2: In online catalog, add notes and full text link
- Step 3: Move PDF to “added” folder

**Results:** More steps, more coordination, higher risk of errors, user frustration

# The Unexpected

---

- Privacy
  - Dean's signature
- Signature Page
  - Dean, advisor, committee members
- Resolution
  - Create blank signature page
    - Adobe Illustrator
  - Replace files in eScholarship



# Estimate vs. Actual Per Title

	<b>Estimate (Minutes)</b>	<b>Actual (Minutes)</b>
<b>Scanning</b>	45	45
<b>Quality Control</b>	45	25
<b>OCR Abstract</b>	20	30
<b>Add to IR</b>	20	10
Signature Page		25
Replace File		5
<b>Project Management</b>	15	30
<b>Total</b>	<b>145</b>	<b>170</b>

Under estimated

Over estimated

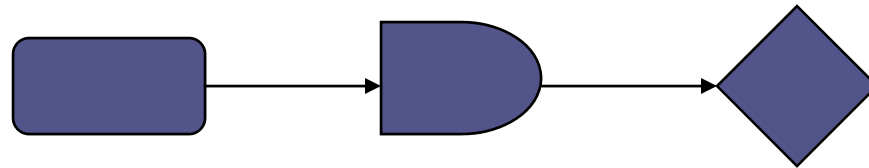
# Estimates vs. Actual Project

	Estimate		Actual	
	Hours	Cost (\$)	Hours	Cost (\$)
<b>Scanning</b>	225	4,500	240	4,800
<b>Quality Control</b>	225	5,625	133	2,926
<b>OCR Abstract</b>	100	2,500	160	3,520
<b>Add to IR</b>	100	2,500	54	1,890
<b>Signature Page</b>			133	2,926
<b>Replace File</b>			26	910
<b>Project Management</b>	75	2,625	160	5,600
<b>Hardware / Software</b>		10,000		990
<b>Total</b>	<b>725</b>	<b>\$27,750</b>	<b>903</b>	<b>\$23,562</b>

# Outsourced Costs

	<b>Actual</b>	
	<b>Hours</b>	<b>Cost (\$)</b>
<b>Quality Control</b>	133	2,926
<b>OCR Abstract</b>	160	3,520
<b>Signature Page</b>	133	2,926
<b>Total</b>		<b>9,372</b>

# Visio in Your Handout



# Long Term Coordination

---

- Between the Library and the Graduate School on the dissertation approval process
- Between Cataloging and Systems

# Project Evaluation

---

- Budget
- Usage statistics as of 10-26-2008
  - 382 dissertations submitted
  - 33,385 full text downloads
  - One 2005 dissertation on dengue fever has been downloaded over 800 times
- Visibility

# Future Directions

---

- Administrative
  - Manage copyright issues
  - Create a marketing and promotion plan
- Self-submission of dissertations
- Content recruitment
  - Specialized student scholar programs
  - Open access journal articles
  - GSBS student publications

# Conclusion

---

- Success factors
  - Library funding, support, management, skills
  - Buying a hosted product
  - Support of Graduate School Dean
  - Small, defined project
- Future success
  - Continued funding
  - Dedicated repository staff
  - Increased faculty and department participation
  - Greater campus awareness



# Thank You!!



**Mary Piorun, MSLS, MBA, AHIP**  
**Associate Director**

**Technology Initiatives and Resource Management**

508-856-2206 – [Mary.Piorun@umassmed.edu](mailto:Mary.Piorun@umassmed.edu)

**Lisa Palmer, MSLS**  
**Catalog Librarian**

508-856-4368 - [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu)

**Presentation slides and handouts:**

[http://escholarship.umassmed.edu/lib\\_articles/96/](http://escholarship.umassmed.edu/lib_articles/96/)

**Detailed journal article about this project:**

[http://escholarship.umassmed.edu/lib\\_articles/94/](http://escholarship.umassmed.edu/lib_articles/94/)