

Lamar Soutter Library

Library Publications and Presentations

University of Massachusetts Medical School

Year 2007

Challenges and Lessons Learned: Moving
From Image Database to Institutional
Repository

Mary E. Piorun*

Lisa A. Palmer†

James F. Comes‡

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

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

‡University of Massachusetts Medical School, James.Comes@umassmed.edu

This paper is posted at eScholarship@UMMS.

http://escholarship.umassmed.edu/lib_articles/67

Title: Challenges and Lessons Learned: Moving From Image Database to Institutional Repository

Mary E. Piorun
Lamar Soutter Library
University of Massachusetts Medical School
55 Lake Avenue North
Worcester, MA 01655 USA
mary.piorun@umassmed.edu

Mary Piorun is Associate Director for Library Systems. In this capacity she manages the Library's technology initiatives and oversees the implementation and management of software/hardware used in the daily operations of the Library. Mary is currently the project manager for the Library's institutional repository initiative.

Lisa A. Palmer
Lamar Soutter Library
University of Massachusetts Medical School
55 Lake Avenue North
Worcester, MA 01655 USA
lisa.palmer@umassmed.edu

Lisa Palmer is Catalog Librarian at the Lamar Soutter Library, UMass Medical School. In this position she oversees the bibliographic and authority control activities for the library catalog, as well as working on the Library teams charged with implementing an institutional repository and redesigning the Library's website.

Jim Comes
Lamar Soutter Library
University of Massachusetts Medical School
55 Lake Avenue North
Worcester, MA 01655 USA
james.comes@umassmed.edu

Jim Comes is Associate Director for Reference, Education and Research Services. Jim oversees the instructional programs, liaison services and reference services to students, faculty and researchers.

Abstract**Category:** Case Study**Purpose:** The purpose of this case study is to chronicle the Library's effort to build an educational image database, and how the project developed into an institutional repository.**Methodology/Approach:** The case is divided into three phases and highlights the organizational, political, technological, and resource issues that are unique to a specialized library with a medium-sized staff, lacking the resources of a traditional university campus. The case concludes with a list of barriers and facilitators to success and a summary of lessons learned.**Findings:** A library with limited staff, funding, and systems development resources can initiate and support an institutional repository. Facilitators of success include clear lines of authority, a strong champion, and the appropriate technology for the project.**Value:** This case serves as an example to libraries that are in the beginning phases of developing an institutional repository by discussing the barriers to and facilitators of success.**Keywords:** Medical libraries, Institutional repositories, Project management

Title: Challenges and Lessons Learned: Moving From Image Database to Institutional Repository

Introduction

The Lamar Soutter Library at the University of Massachusetts Medical School was founded in 1970. It is considered a medium-sized academic health sciences library with just under 300,000 volumes. The library staff consists of 42 FTE, serving 950 students and 1500 faculty and researchers.

In the spring of 2003, the Library was actively seeking an opportunity to build an image database. The effort was led by the Library Director, who was in communication with both faculty members and leaders in the school's Information Services (IS) department. Through these connections the Library learned about a small database of histology images used for teaching medical students. The database was maintained by a faculty member in the Cell Biology department. Cell Biology asked IS to help maintain the database, provide storage, and expand the availability of the database to all members of the department, and possibly other medical school faculty.

IS was hesitant to take on the level of support requested, viewing the project as benefiting only one department. At this point the Library became involved. The Library Director presented the Library as a willing collaborator with the skills necessary to move the image database from a departmental project to a campus-wide project.

Phase I: Getting Started

The Library assembled a small team with representatives from its systems, technical services, and reference departments to work with IS and Cell Biology. The Library staff had extensive experience working in teams, but this was the first time the Library would participate in a formal team environment outside the Library.

Each member brought a unique skill to the team and, more importantly, each department played an important role. The Library contributed skills in indexing, search and retrieval, copyright, and project management. The Library had also established positive working relationships with many of the academic departments. IS provided technical and financial support, and Cell Biology supplied content. An estimated 500 images were stored in the Cell Biology database.

Ideally, project goals are formulated first, and then funding and staffing of the project are added. The failure to follow this standard approach proved to be a drawback to the long-term success of this phase. The project goals -- buy software, train users, support teaching -- were high-level, action-oriented, and not linked to the strategic objectives of the Medical School. Most importantly the team members lacked a shared vision. The Library's vision was to develop a database that could be used by all faculty for teaching purposes; translating this into formal goals and objectives would have taken too long from the perspective of the other institutional players. For example, IS made end-of-year money available and their priority was to make a quick software purchase.

The Cell Biology database was built with a program called Cumulus. Cumulus is a digital asset management (DAM) program, with a large user community based in the photography and multimedia industries. There was significant pressure from IS and Cell Biology to continue

using the Cumulus platform. With very little discussion of options IS determined that the project would go forward by upgrading from an individual license of Cumulus to the enterprise level. The server software was loaded onto the only available machine, a ten-year-old server without a maintenance contract.

Funding was made available for training to familiarize team members with the Cumulus software. During the training session, the team realized Cumulus was available in multiple versions. IS had purchased a version that restricted the number of characters in a search string. The Library considered this to be a major flaw in the software. The team also learned a Java programmer would be required to make any changes to the web interface. In addition, the server had serious performance issues and would need to be replaced.

Not to be deterred, the team moved forward with the database project. The team focused on setting policies, contacting and training faculty, and moving the images from Cell Biology into the shared version of Cumulus. At this point it became evident that not all of the departments in this project had the same goals. Library team members were intent on demonstrating the value that librarians could bring to this project and were not attuned to the goals of other team members. For example, IS wished to limit their involvement to technical support; they were unwilling to commit resources to training, copyright, and marketing. The team also noticed that Cell Biology was not very enthusiastic about sharing their images across the institution.

As the Library proceeded to train faculty and solicit additional images, unanticipated technical, and policy issues emerged. For example, faculty affiliated with the Medical School's clinical partner were on a separate network and unable to reach the Cumulus server to post new images. IS worked to find a solution, but did not have the resources to follow through. A policy issue regarding acceptable formats surfaced because Cumulus allowed video to be posted, but did not have streaming ability. One faculty member posted an hour-long video that took 20-30 minutes to download before it could be viewed. This left users with a less than satisfactory first impression.

The team was intent on making the Cumulus system work as a tool for searching, retrieval, and preservation of images and documents. Two specific issues required the team's attention. First, the default web interface was not acceptable. It could only be used for searching; many of the customized fields added to the metadata structure were not indexed, searchable, or viewable. Making changes to the display required Java programming skills. Second, utilization of the client interface would require the coordination of a mass deployment to all school desktops, along with user training. This point is important, because new images could be entered only through the client interface.

The team had been working together for over a year. During this time, Cell Biology experienced some staffing changes and retirements of faculty who originally championed this project. As a result, representation at weekly meetings ceased, and the team discovered that the 100 images Cell Biology had contributed to the shared database had been removed. IS was still involved in the project, but because the server was up and running, their commitment to overall project planning and deployment was dwindling. They did not have the resources to provide a Java programmer to work on customizing the web interface.

The team meetings started to focus on the limitations of the Cumulus software. Members felt they had taken the software program as far as they could. To get the team back on track, a half-day "retreat" was organized. This gave the team an opportunity to review the work that had been completed to date and plan for the next phase.

Phase II: The Transition: New Directions and Evaluation

The team unanimously agreed that the Cumulus system would not meet the needs of the Medical School. Most team members wanted to explore solutions such as institutional repository (IR) software, which might be more useful and successful than an instructional image database alone. The benefits of an IR had been discussed for several years in the literature (Crow, 2002; Gibbons, 2004). Particularly compelling to the team were: 1) establishing a showcase for the school's research, teaching, and scholarship, including unpublished dissertations; 2) providing open access to research; and 3) developing a user-friendly process for faculty, researchers, and staff to promote and distribute their work.

In August 2004 the Library Director recognized that the current project with IS was limited by their budget, staffing, and technology preferences. The Library Director and IS agreed that the team would reorganize to form a joint IS/Library task force to investigate IR software, with the Library taking the lead role. Library task force members were pleased with this change.

Specifically, the task force was charged with investigating IR products and making a recommendation to the two department heads for a depository system. The task force was expanded to include an experienced IS researcher with close ties to faculty and an interest in libraries, database systems, and informatics.

The new focus and enhanced composition invigorated the task force, and members quickly began researching the products available. A table comparing technical features and system functionality (Open Society Institute, 2004) was especially helpful.

In two months, the task force produced a preliminary report with four major recommendations:

1. Continue with the Cumulus system on a limited basis, while preparing for conversion to a new system in the future;
2. Perform an assessment to identify needs, raise awareness, and generate support for establishing an IR;
3. Develop a demonstration project to build institutional support and give the task force experience with building such a repository;
4. Assess the level of institutional support required for sustaining an IR.

The recommendations were approved and the task force was given one year to implement these goals. The team quickly decided to upgrade the Cumulus software in order to retain vendor customer support, and purchased a new server. Plans were made to create a display in the Library to raise awareness about institutional repositories. The Library offered to host a local ACRL Information Technology Group meeting about IRs.

At this time, what appeared to be a golden opportunity was presented to the Library Director. An office in the health care consulting division of the medical school asked for a proposal to create a centralized digital repository of their publications. The documents, created for consumers and health care providers, were in several languages and stored in multiple formats (paper and electronic) in various locations. This project would be a way to establish new relationships with a growing organization, one that could help with funding.

Library staff on the team wrote a detailed proposal for a one-year pilot project that included an implementation plan, staffing and budget estimates, and a projected timeline. This proposal, with a \$200,000 budget, was submitted in February 2005. The budget included staff, hardware, software programmers, office equipment, and office reconfiguration; everything the team thought was needed to be successful after the difficult first phase of the project. However, funding was not available and the proposal was abandoned. Library staff acquired valuable experience with proposal planning and writing. Since the Library pursued this venture independently of IS, the task force was splintered and the IS members ultimately dropped off.

In July 2005 the license for the Cumulus software came up for renewal. The Library and IS agreed it was time to retire the Cumulus system, under the condition that the remaining assets and metadata would be preserved and migrated to a future system, if desired. The team was interested in pursuing a demonstration project with a vendor product, and the renewal funds could be useful.

By early fall 2005 the team had a list of three IR software products under consideration: 1) DSpace, an open source product developed by MIT and Hewlett-Packard; 2) Digital Commons from ProQuest, a hosted product based on the bepress technology; and 3) Open Repository, a hosted service based on the DSpace technology from BioMed Central in the UK.

A “score card” was created to rate and evaluate the three products. Five criteria were assigned weights (Table I).

Take in Table I

The Tools and Administration areas were weighted most heavily because the Medical School is a separate campus physically and administratively within the larger university system. The Library does not have access to typical on-campus system development resources, such as engineering or computer science departments and work-study students. Based on past experience, the team was keenly aware that any technical skills needed would have to be provided by Library staff, and weighted the corresponding areas accordingly. It was critical that the product be robust, require little special programming, and be implemented and maintained with current Library staff.

In order to make these judgments, team members read articles, monitored discussion lists, reviewed library websites, talked with users from other libraries, attended various workshops, and saw product demonstrations. A spreadsheet summarizing the cost of running open source IRs proved to be a helpful resource for estimating costs (Kemp, 2005).

Each team member scored the three products individually, keeping notes on their thoughts and impressions. The group came together for a joint discussion and to determine a final score.

For a second time an outside “opportunity” sidetracked the team. The university system’s President’s Office asked the Library for assistance in creating a searchable, digital archive of the University’s Board of Trustees meeting minutes, dating back to 1863. The Library was looking for a demonstration project, and this seemed like the ideal project to take on. It was a manageable size, and the documents were in a variety of formats: some handwritten, some typed, and some electronic. The team was excited that the Library – one of five in the university system – had been approached by the President’s Office. Again, a subset of team members began to work on a project proposal, which was submitted in November 2005, with a budget of \$100,000. Submitting this proposal raised the profile of the Library system-wide.

Unfortunately, funding proved to be the stumbling block. The staff at the President's Office were impressed with the thoughtfulness and level of detail in the proposal, but did not have the financial resources required to digitize the meeting minutes.

Library staff put the disappointment behind them and went back to making a final decision on an IR software product. In the end, the Library decided to purchase a two-year license for ProQuest Digital Commons. The most important factors in the decision were:

- A recognition that the Library did not have the resources to support an open source product such as DSpace;
- Open Repository was too new and untested;
- Digital Commons included functionality that would make it easier to promote and populate the repository, including e-mail alerts and faculty researcher pages;
- Positive references from other librarians about Digital Commons.

Phase III: Getting Started - Again

The team licensed Digital Commons in January 2006 and named the school's repository "eScholarship". After initial testing with Library staff publications, the team quickly focused on two projects: 1) dissertations from the Graduate School of Biomedical Sciences (GSBS), and 2) processing an XML data load from the ProQuest databases and PubMed. The team's intent with these projects was to populate the repository quickly, generate visibility, and gain support across the institution.

Project 1: Dissertations

The GSBS was founded in 1979 and has produced over 300 doctoral dissertations, very few of which were submitted in electronic format. The Dean was interested in having each dissertation posted to the web and fully searchable. GSBS dissertations had never been sent to UMI and thus could not be automatically loaded by ProQuest into our repository. The Library did not have any experience in managing a large-scale scanning project. A brief report was written for the Library Director outlining the pros and cons of either outsourcing the project or handling the project in-house. The Director decided to fund the project in-house, ensuring control over the materials and quality.

Before the team could proceed, questions about copyright permissions needed to be addressed. It was decided that each alumnus would be asked to sign a form granting permission to convert their dissertation to an electronic format and to add it to the repository. The team worked with GSBS to develop this form and obtained a list of names and addresses of their alumni. Future graduates will be given the option to delay adding their dissertation to the database for one year, if the research is ongoing or pending publication.

The team's systems representative led efforts to 1) contact alumni to obtain and track permissions, 2) manage the scanning and OCR process, and 3) work with ProQuest to obtain the level of customization needed for this section of the system. The catalogers on the team led efforts to 1) define metadata requirements, 2) identify ways to reuse the bibliographic data from the Library's OPAC, and 3) develop a workflow to add the electronic files to the eScholarship database. To date, permission has been obtained from 150 of the 300 alumni.

Project 2: XML Data Load

The team immediately asked the vendor to perform an XML data load from the PubMed database. This functionality would harvest metadata -- citation, abstract, medical subject headings -- for the institution's faculty publications. In medicine, PubMed is by far the most popular database, indexing most of the journals in which school faculty are published. This harvest provided the team with a core piece of content (6,000 citations) that could be used to jumpstart the IR project. From promotional materials, the team thought the data loader was a mainstream technology, but it quickly became evident that no other site was using this functionality. The team found for the first time in three years they were on the "bleeding edge" of an untested technology in the library environment (Endry, 2003).

It was now the responsibility of the team to test unproven workflows and identify "bugs" in the system. There were increased calls to technical support, in many cases just to learn how the system functioned so a workable process could be developed. Each time the team thought they had a possible solution to a problem, a new obstacle emerged.

At this point, it is not clear that the Library will be able to process the 6,000 citations that were harvested. The team continues to weigh the pros and cons of utilizing the harvested metadata, versus developing a plan to work directly with faculty members to deposit their recent publications.

The dissertation and data load projects have raised many challenges for the team, such as: bringing value and recognition to a resource that is still in development, creating a culture of self-archiving, and managing the increased workload.

In order to meet these challenges, the team has set the following priorities: continue with the two projects that are underway, document procedures, develop policies, create a marketing and promotion plan, seek new opportunities to recruit content, work directly with faculty members to generate support, and manage copyright issues.

Future projects include working with specialized student scholar groups, reaching out to the Graduate School of Nursing to digitize their dissertations, and working with the Women's Faculty Committee to help highlight the works and research of women faculty members.

Discussion

This paper describes the evolution of a digitization project from an image database created in Cumulus to a growing institutional repository hosted by Digital Commons. The case study illustrates the barriers to and facilitators of success to consider when embarking on such an endeavor. By identifying these barriers and facilitators, we can then infer a number of lessons learned.

Table II compares the barriers to success for the two projects. Clearly, there were more barriers to the image database project than the IR project. Barriers for the image database included funding, project goals and lack of a shared vision, unclear decision making, inadequate technical support, and finally, an unclear understanding about what the software could really do. One department's pressure to spend end-of-year money drove the decision-making process. Although involving several cross-campus departments in the project initially seemed positive, this led to problems as the project progressed. Barriers to success of the IR include: limited faculty involvement in the project, only four Library staff devoted to the project, increased workload, recruiting content, and creating workarounds for software limitations. Digitizing

dissertations and obtaining permissions are time-consuming processes, as are many of the tasks associated with running an IR. The barriers to success for the image database were organizational and political, while the barriers to success for the IR are process-related.

Take in Table II

Table III illustrates the facilitators of success for the image database and IR projects. Facilitators of success for the image database included having the involvement upfront of faculty who were willing to supply content. Team members' efforts to keep the project moving forward also contributed to the limited success of the project. Clearly there are more facilitators to success for the IR project. The team learned from the political and organizational problems that occurred in the image database project. It was clear from the outset that the Library would fund, support and manage the project. The decision to buy a hosted product allowed the Library to implement the IR independently. In addition, the IR had the support of a school administrator outside the library (GSBS Dean).

Take in Table III

Lessons Learned and Conclusions

The barriers and facilitators to success have led the team to infer a number of lessons learned that may be helpful to others considering an IR project.

1. *Teamwork:* Building an institutional repository requires the skills and knowledge of a wide array of Library staff. Include staff with skills in project management, outreach, technology, and metadata. Build a team that can be flexible and grow with the project.
2. *Reflect:* Set aside time periodically to reflect on the work that has been accomplished. Celebrate victories and acknowledge failures. It will keep the project moving forward and stop the team from repeating mistakes.
3. *Time commitment:* Building a repository takes time. Balance the success of the repository with the amount of additional work that library staff can take on. Project leaders need to make sure everyone stays involved and the workload is balanced.
4. *Goals:* Set a variety of goals: short-term, long-term, attainable, and stretch goals. Remember, each team member will also have personal goals that they are working towards.
5. *Be creative:* There are many hurdles to overcome when building an institutional repository. Libraries with limited resources need to think creatively and be persistent.
6. *Getting sidetracked:* Opportunities will come along that are too good to pass up. Sometimes these are costly and disrupt the flow of the team. Ask if there is something to be learned, as well as gained, as the team pursues these opportunities.
7. *Perfection:* Don't allow the team to become paralyzed. Avoid the temptation to procrastinate, or even stop, because team members don't have all the answers or are afraid of failing.

8. *Champion*: Build relationships with key decision makers who will champion the project. Look for win-win opportunities within the Library and across the institution.

The team continues to ask, "How will we know if we are successful?" Two simple measures used by many sites (Bailey, 2006) are 1) the number of documents in the repository (264 for May – September, 2006) and 2) the number of downloads (1,679). The team is pleased with its progress on the eScholarship project, particularly the dissertations effort. One 2005 dissertation alone has been downloaded almost 130 times in 3 months.

For the team, true success will be measured by continued funding, the hiring of new staff dedicated to the repository, increased faculty participation, and greater campus awareness of the value of a scholarly repository.

Acknowledgement: The authors gratefully acknowledge the support and contributions of Elaine Martin in the writing of this article and throughout this project.

References

Bailey, C. (2006), *SPEC Kit 292: Institutional Repositories*, Association of Research Libraries, Washington, DC.

Crow, R. (2002), "The case for institutional repositories: a SPARC position paper", available at: <http://www.arl.org/sparc/IR/ir.html>

Endry, M. (2003), "Should a company be on the bleeding-, leading-, or following-edge of technology", in Laube, D. and Zammuto, R. (Eds.), *Business Driven Information Technology*, Stanford University Press, CA, pp. 21-25.

Gibbons, S. (2004), "Establishing an institutional repository", *Library Technology Reports*, Vol 40 No 4

Kemp, R. (2005), "Listserv summary of the costs of running an open source repository", available at: <http://library.uncw.edu/web/faculty/kempr/documents/listserv-summary-IR-open-source-costs.xls>

Open Society Institute (2004), "A guide to institutional repository software", 3rd edition, available at: <http://www.soros.org/openaccess/software/>

Table I: Score Card for Evaluating Repository Software

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

Table II: Barriers to Success

Image Database (Cumulus)	Institutional Repository (Digital Commons)
Conditions placed by funding source	Lack of faculty involvement
Unclear project goals and shared vision	Limited number of staff devoted to project
Not the right mix of team members	Workload
Unclear line of authority	Content not available upfront
Unclear decision-making process	Software did not meet expectations
Inadequate technical support	
Software did not meet expectations	

Table III: Facilitators of Success

Image Database (Cumulus)	Institutional Repository (Digital Commons)
Faculty involvement	Clear decision making
Content available upfront	Clear line of authority over the project
Persistence of team members	Support of GSBS Dean
	Funded by Library budget
	Minimal technology support needed locally