



Interim Reports
submitted to the
Institute for Museum and Library Services
2003-2005

Joanne Grossman, Project Director

Doors of Opportunity: The Women Physicians Digitization Project

Interim Narrative Performance Report (First)

In this report we review the status of project goals and objectives after six months (October 1, 2003 – April 1, 2004). Unusual circumstances have slowed down the expected pace of work somewhat: The University's affiliated for-profit hospital system (Tenet Healthcare, Inc.) leases space to University departments, including the Archives and Special Collections on Women in Medicine. Last Fall Tenet decided to close the building in which the Archives is housed, though it was unclear as to when this would take place. Considerable staff time has been spent determining an appropriate new space for the Archives, impeding progress on the digitization project. To date, new space for the Archives is still undetermined, although now it seems that the move will take place in approximately three months.

The Digital Project Archivist joined the project staff on January 5, 2004, contributing to the delay. With a move imminent, the first priority became pulling the materials to be digitized and organizing them separately. The Assistant Archivist position (part of the University's cost-share), which had been vacant, was filled on March 15, 2004. The Metadata Specialist and Scanning Technician will be hired shortly.

Below are specific goals and their status:

A. Outcome Evaluation

Staff attended the "Outcome-Based Evaluation for Preservation & Digitization" workshop in Washington, D.C. (December 11 and 12, 2003). This training prepared staff to build and manage an appropriate outcome evaluation for this project.

B. Exploration and Choice of Open-Source Software

The decision was made to utilize an existing open-source software rather than designing and building a database internally. Project staff reviewed three potential open-source options, DSpace, FEDORA, and Greenstone Digital Library System (GSDL), ultimately choosing GSDL for deployment. The software has been installed, including its Digital Librarian Interface and staff is beginning to customize it as needed to support this project.

C. Digital Preservation

Storage Resource Broker (SRB) will be utilized with this project as a mechanism to support long-term preservation. The system, developed by Reagan Moore, Ph.D., and the Super Computing Center at the University of California at San Diego, utilizes data grids to create persistent archives, such as the digital archives developed by the NSF's National Science and Digital Library (NSDL) projects. Staff plan to propose that the digital archives created by this women in medicine project will be searchable along with the results of this larger effort. Dr. Moore will visit Drexel and consult with project staff later this year.

D. Collection Sub-set

Materials intended for digitization vary in format and reside in numerous collections. In preparation for scanning and cataloging, project staff have begun creating a sub-set of collection items to be included in the project. These materials are carefully housed and identified, and will be easily refiled within the original collections from which they were removed.

E. Copyright

Project staff have met with both a University copyright attorney and expert outside counsel to prepare digitization guidelines related to copyright. They will continue to meet throughout the project. Existing resources regarding copyright, rights to publicity and rights to privacy also act as reference guides in determining the University's right to publish Archives materials online. Staff have begun making final determinations in preparation for digitization.

F. Selection Committee

This group of stakeholders met on February 9, 2004 and shared ideas regarding project content and its digital presentation to users, particularly high school and undergraduate students. Two additional meetings are planned to work toward developing appropriate curriculum material for these target audiences.

G. Project Website

The project website is currently under development at <<http://med.library.drexel.edu/archives/libtemplate1.js>>. The site currently describes the project and will, over time, include project standards and documentation as well as IMLS Interim Reports. We will maintain this information as part of the end product.

H. Equipment Purchase

Software and hardware needs were re-evaluated and appropriate equipment was purchased.

Doors of Opportunity: The Women Physicians Digitization Project

Interim Narrative Performance Report (Second)

This report reviews the status of project goals and objectives after the second six month period of the project (April 1 – October 1, 2004). As anticipated in the first report, the Archives and Special Collections on Women in Medicine did indeed move to other sites on campus, slowing the expected pace of project work. Additionally, internal organizational changes affected the Archives and the reporting structure for the digitization project. These resulted in an IMLS-approved change to a new Principal Investigator and submission of a request for a no-cost extension.

Below are specific goals and their status:

A. Project Staff

Internal changes have moved the Archives department from the University libraries to the Institute for Women's Health and Leadership, which is part of the College of Medicine. This brings the Archives back to its place of origin, before Drexel University absorbed the institution that is now the College. One of the results of this reorganization is a change in Principal Investigator. The new P.I. is Arnold Smolen, Associate Dean for Information Technology at the College of Medicine. Dr. Smolen is well-versed in the College's unique and significant role in the history of women physicians. His understanding of the requirements for digitization, and of this project in particular, have kept the project on schedule and moving smoothly through this period of physical and organizational transition.

The Digital Project Archivist position, held by Margaret Graham and originally budgeted as full-time, is now established at .7 FTE. Accordingly, the position of Metadata Specialist was expanded to include a greater range of responsibilities and the title changed to Assistant Digital Project Archivist.

A Drexel library staff member, Systems Specialist Tom McLaughlin, was removed from the project and replaced by Network Coordinator, Ian Richmond. These internal changes influenced the hiring decision for the position of Assistant Digital Project Archivist, favoring a candidate with strong technical skills. Charles Dennis started in this position on May 17, 2003. As project needs evolved, the title was changed to Digital Development Archivist, reflecting the position's expanded scope.

The Scanning Technician, Mike Ratti, joined the project team on September 7, 2004.

B. Digital library system

As noted in the first report, the decision was made to utilize existing open-source software rather than designing and building a custom database internally. Project staff chose the Greenstone Digital Library System (GSDL). The software was installed, including its Digital Librarian Interface, and staff began to customize it as needed to support the project. Conflicts in the GSDL system, in combination with a shift in systems support as a result of organizational restructuring, resulted in a decision to design a MySQL relational database rather than rely on GSDL.

As GSDL was being explored, project staff designed and built an XML and Perl based flat file system for tracking project materials and their metadata, dubbed dbTrack. The system was initially designed to handle the minimal amount of metadata necessary for scanning, and a script was written to export that metadata into GSDL. From there, all descriptive metadata would be entered through Greenstone's Digital Librarian Interface. Running into problems with the customized metadata set in GSDL, dbTrack was expanded to handle all the metadata, which could still be easily exported into Greenstone, bypassing the Librarian Interface.

When the decision was made to move away from Greenstone and build a custom database, dbTrack was already in place as the project's metadata repository. While the new database is in development, production continues, with all metadata being written to dbTrack. When the new system is complete, all dbTrack data will be exported to the relational database. Ultimately, the relational database will serve as both a production database and means for online delivery of the digital materials.

C. Cataloging

Approximately 400 collection-level MARC records have been created and entered into the University OPAC, providing links and context for item-level digital objects. While some subject cataloging remains to be completed, all the items currently selected for the project have full records in the OPAC. Item level records continue to be created with over 1600 items partially catalogued and 50 records completed.

D. Collection Sub-set

Materials intended for digitization vary in format and reside in numerous collections. In preparation for scanning and completion of cataloging, project staff have created a sub-set of collection items to be included in the project. These materials are carefully housed and identified, and will be easily refiled within the original collections from which they were removed. To date, more than 1600 items (21,000 pages) have been identified. Preliminary item-level records have been created for each.

E. Copyright

Copyright guidelines related to the project have been established. Project staff continue to meet with both a University copyright attorney and expert outside counsel to clarify any issues.

Existing resources regarding copyright, rights to publicity and rights to privacy also act as reference guides in determining the University's right to publish Archives' materials online. A Copyright Office researcher has also provided necessary information in making final determinations in preparation for digitization.

F. Selection Committee

The Selection Committee met on February 9 and discussed a number of issues with great commitment and enthusiasm. These stakeholders, both internal and external, considered the implications of including specific materials given the target audience of high school students. Concerns regarding images of dissected bodies, negative stereotypes found in textual material, items including controversial medical issues, etc. were expressed and resolutions agreed upon.

One of the results of this meeting was a decision to meet with a group of project staff and educators to identify specific user needs of high school and undergraduate students. A meeting on October 6 will address issues regarding high school curricula, outreach methods, the project web interface design and usability testing.

G. Digital Preservation

Early work on digital preservation resulted in a decision to use Storage Resource Broker (SRB) as a mechanism to support long-term preservation. A different approach was chosen in consult with Michael Lesk, Professor in the Library and Information Science Department at Rutgers University and an expert in digital preservation: Standard, non-proprietary formats for information will be utilized. Maintenance of current files will be in accordance with normal computer center techniques, including daily backups of disks onto magnetic tapes.

Dr. Lesk recommended that data be backed up to external hard drives and deposited with outside institutions in the event that the project is lost locally. Two institutions with an interest in the collection, such as the National Library of Medicine and the Arthur and Elizabeth Schlesinger Library on the History of Women in America at Harvard University, will be contacted as possible repositories for back-up data. (The full digital preservation report can be viewed at http://med.library.drexel.edu/archives/digproj/pres_report.html)

Additionally, the project team has decided to meet with a preservation consultant twice more throughout the project.

H. Equipment Purchase

Additional software (e.g. OCR) and hardware (e.g. external hard drives) were purchased as needed.

I. Project Website

The project website is located at <http://med.library.drexel.edu/archives/digproj/index.html>. The site currently describes the project and includes project documentation, the preservation consultant's report, and IMLS Interim Reports. We will maintain this information as part of the end product.

Women Physicians: Presenting Their Past Digitization Project

Interim Narrative Performance Report (Third)

This report reviews the status of project goals and objectives after the third six month period of the project (October 1 – April 1, 2005). As detailed in the last report, the Archives and Special Collections on Women in Medicine moved to various sites on campus last June. Below are specific goals and their status:

A. Project Staff

As described in the previous report, the project now has a new Principal Investigator. Dr. Arnold Smolen has led the project team through a complex transition, offering support and institutional authority to keep the project on track. As part of internal institutional changes, the project was moved from a library server to a College of Medicine server, with Dr. Smolen guiding the process and championing the needs of the project. The Archives, which is now a core program of the Institute for Women's Health and Leadership, has experienced support, interest and solid leadership in all areas, including the digitization project.

Michael Ratti was mentioned in the last report as having started in early September. Mike has taken the lead in scanning, and has also brought significant cataloging skills to the position.

A student worker, Kerry Corrigan, was hired in the fall and continues to contribute to the project with her scanning skills. Mustanser Bader is a work/study student who efficiently handles scanning tasks as well. Both of these team members work one day per week.

Digital library system

In preparation for the development of the publicly accessible digital library system, the production tool already in place, dbTrack, was converted from an XML/Perl-based system to a relational database system built on MySQL. The new database design supports both the production tools and online public access. dbTrack was completed in December 2004, and the public side of the system, dubbed xDL, will be ready for user testing in May.

The project now lives on a server housed by the College of Medicine, dedicated to the Archives department and robust enough to manage the project needs. Due to institutional changes, the project moved several times before finding this home. The changes have shifted the responsibilities of the Digital Development Archivist to nearly 100% of his time being spent on development and systems administration. Consequently, additional project management and cataloguing responsibilities have been taken on by other project staff. Development remains on schedule.

xDL's access points include browsing by project categories, subjects, names, dates and format. The system offers Boolean supported basic and advanced search and a keyword

feature that searches all fields as well as full text. In the advanced search, users have the option of limiting their search by date, format or collection.

Individual items can be viewed page by page, by series or by all items available in a particular collection. Additional functionality includes a portfolio system, whereby users can collect and review items of interest and/or request additional information about those items.

Several workflow steps have been automated, written in Perl and run from the command line, including the creation of delivery jpegs, uncorrected OCR files and watermarking of images. OCR files requiring correction (where the text file will be displayed) will be processed at desktop machines. Automating these processes has streamlined our workflow and freed staff time to complete more cataloguing and scanning.

Image quality control is done by project staff. As jpegs are generated, they are reviewed for sharpness, skew and tonal accuracy. Due to the implementation of the new server, this process has been somewhat delayed, but no long term problems are anticipated.

Dedicated University Relations design staff are currently developing the interface design. It will be completed and implemented in time for user testing, planned for mid-May.

Testing of the digital library system is scheduled for mid-May. The process will test five users from each of our targeted user groups: high school students; undergraduate students; and scholars and independent researchers, including genealogists. We anticipate conducting both on-site and remote testing. Project staff is currently designing the tasks and queries that will assess the usability of the digital library system.

C. Project materials

To date, more than 1700 items (25,000 pages) have been identified for the project. 850 items have completed item-level records.

The project Scanning Technician, student worker and work/study staff have scanned 10,000 pages of material, adhering to standards assigned.

One out of a total of 12 oversize or bound items have been sent to an outside vendor where an overhead scanner can accommodate these collection pieces. Currently, twenty-six items have been disbound for safe and thorough scanning and will either be rebound or rehoused before they are reintegrated in the larger collection.

Project staff continues to meet with both a University copyright attorney and expert outside counsel to clarify any issues. A Copyright Office researcher continues to provide information to assist in the process of final determinations for individual objects. All but 94 of the 1700 items identified have been approved for online publication.

D. Educators' Committee

The Selection Committee met in February, 2004 and discussed a number of issues with great commitment and enthusiasm. One of the results of this meeting was a decision to convene a group of project staff and educators to identify specific user needs of high school and undergraduate students. A meeting in October, 2004 addressed issues regarding high school curricula, outreach methods, the project web interface design and usability testing. The results were very positive, including identification of potential usability testing groups who will help shape the functionality and design of the site. High school and undergraduate educators offered connections to various associations and organizations in order to disseminate information about the project and collect input on curricula needs. Project staff will join a newly formed Philadelphia-area group of educators and archivists who will address technology issues for students.

E. Digital Preservation

Last summer the project team met with Michael Lesk, Professor in the Library and Information Science Department at Rutgers University. Additionally, the team is planning to meet with another consultant in late spring to secure additional advice on this crucial issue for the project and Drexel's long-term commitment to digitization.

F. Dissemination

Our primary goals in disseminating information about the project are 1) to share our experience with others who might learn from us, and 2) increase exposure to the digital library and expand our user base. Preliminary arrangements have been made to speak about the project at regional professional meetings in Fall, 2005, including the Mid-Atlantic Archivists Conference (MARAC) and the Delaware Valley Archivists' Group. Other fall meetings under consideration include the Association of American Medical Colleges (AAMC). We are also submitting abstracts for meetings in 2006, including a joint meeting of the American Association for the History of Medicine (AAHM) and Archivists and Librarians in the History of the Health Sciences and the Society of American Archivists.

Women Physicians: Presenting Their Past Digitization Project

Interim Narrative Performance Report (Fourth)

This report reviews the status of project goals and objectives after the fourth six month period of the project (April 1, 2005 – October 1, 2005). Below are specific goals and their status:

A. Project Staff

The project continues to thrive under Dr. Arnold Smolen's leadership, and the Archives, a core program of the Institute for Women's Health and Leadership since August, 2004, continues to experience support, enthusiasm and solid leadership in all areas, including the digitization project.

A student worker, Kerry Corrigan, was hired last fall and contributed to the project with her scanning skills through the early spring. Laura Stroffolino joined us in early summer and continues scanning. Laura also edits catalog records and has begun transcription.

B. Digital Library System and User Testing

In May 2005, the digital library system was ready for user testing. A temporary version of the interface was in place, essentially a rough draft with representative visual elements and 80% functionality. Testing was conducted with representative user groups: high school, graduate and medical students, scholars and genealogists. Eight users were tested on-site using direct observation with verbal feedback and four users ran the test independently from remote sites, providing written feedback. While overall feedback was very positive, there were several common experiences that influenced the final interface design. Most significantly, many users did not easily see how to browse the materials. Another important response was that the subject categories needed more visibility. An outside designer was contracted to supplement internal staff's visual design of the interface and these and test results were addressed in the redesign.

In September 2005, development of the digital library system and public interface was completed. The system supports browsing by subject, names, dates, format and collection, and basic and advanced searching, including keyword search of full text items. Advanced search provides the option for the user to choose subjects or names directly from the project-developed controlled vocabulary and name authority list. Each item is navigable by page, contents and series.

The internally developed search engine, written in PHP, searches all record metadata, as well as searching full text for the 40 percent of the digital collection that is composed of printed text. The search engine takes full advantage of MySQL's Boolean search capability. A portion of handwritten items are being transcribed and will also be full-text searchable. Server-side automated processes are largely complete, including programs that perform OCR

image processing, indexing, pdf output, metadata creation and various other administrative processes.

The system also sports a new image viewer. The viewer is Flash-based and is built on the application, Zoomifier, to deliver jpegs that can be zoomed and rotated. A printable jpeg is also available to users.

Project records are ready to export to the University Library OAI-PMH server. The server is being re-built and will be ready for records by December, allowing our metadata to be harvested by other systems.

With development complete, staff is concentrating on image conversion, cataloguing, quality control and project documentation.

C. Project Materials

To date, more than 1700 items (26,500 pages) have been identified for the project. 1585 items have completed item-level records.

The project Scanning Technician and interns have scanned more than 20,800 pages of material, adhering to standards assigned.

Materials that could not be scanned on project flatbed scanners were sent to two outside vendors, including one oversize bound volume and six printed bound volumes. In addition to the 26 items that were disbound for in-house scanning, 15 – 20 more items will be disbound for in-house scanning.

Project staff met regularly with both a University copyright attorney and expert outside counsel to clarify any issues. A U.S. Copyright Office researcher provided information to assist in the process of final determinations for individual objects. Most identified items were cleared for online publication, and others were selected to replace those that were not available.

D. Digital Preservation

The project team met with Michael Lesk, Professor in the Library and Information Science Department at Rutgers University in July, 2004. Recognizing the need to secure additional advice on this crucial issue for Drexel's long-term commitment to digitization, we are co-sponsoring a forum on Digital Preservation in Cultural Institutions with Drexel's College of Information Science and Technology. On October 28, a panel of three speakers will address this open forum: Anne Kenney, Associate University Librarian at Cornell University Library will address the organizational framework necessary for large-scale preservation; Kenneth Thibodeau, Director, Electronic Records Archives Program, National Archives and Records Administration will focus on preservation of born digital objects, concentrating on technical

issues from a management perspective; Howard Besser, Professor and Director of the Moving Image Archive and Preservation Program at NYU's Tisch School of the Arts will discuss preservation of non-traditional materials such as moving images and art.

The forum has been widely advertised, and while a regional audience is expected, interest has been expressed from all over the country. We plan to videotape the forum and make it available via webcast.

F. Dissemination

Our primary goals in disseminating information about the project are 1) to share our experience with others who might learn from us, and 2) increase exposure to the digital library and expand our user base. We presented at the Mid-Atlantic Archivists Conference (MARAC) on October 22, and will present at the Delaware Valley Archivists Group (DVAG) on November 15. Our project has been included in the Innovations in Medicine and Women in Medicine poster sessions at the Association of American Medical Colleges (AAMC), to take place on November 5-7. We are also submitting abstracts for meetings in 2006, including a joint meeting of the American Association for the History of Medicine (AAHM) and Archivists and Librarians in the History of the Health Sciences and the Society of American Archivists.