

University of Maryland Early Dissertations for Doctor of Medicine (1813-1889): Challenges and Rewards of a Digitization Project

María M. Pinkas, Abra Schnur, Megan Wolff, Sarah Hovde & Carol Harling-Henry

Authors' affiliation: Health Sciences and Human Services Library, University of Maryland, Baltimore, Maryland, USA

Address correspondence to María M. Pinkas, Health Sciences and Human Services Library, University of Maryland, 601 West Lombard Street, Baltimore, MD 21201. E-mail: mpinkas@hshsl.umaryland.edu

Abstract

The Health Sciences and Human Services Library holds in its Historical Collections dissertations required for obtaining the degree of Doctor of Medicine at the University of Maryland during the nineteenth century. A project to digitize these dissertations for inclusion in the UM Digital Archive began in 2011, thanks to an award from the National Network of Libraries of Medicine, Southeastern/Atlantic Region. This article presents challenges encountered, as well as rewards and learning experiences gained during this digitization project, especially in the assignment of accurate metadata, and in the assessment of scanning quality.

Keywords

Digital preservation and dissemination, digital archives, hidden collections, manuscripts, dissertations, metadata management, scanning quality control, PDFs

Introduction

Digital preservation in academia is a rewarding and necessary endeavor in the 21st century. The ability to create and disseminate digital information through institutional repositories and digital archives, such as the Internet Archive, has opened otherwise hidden collections to the general public. This resurgence in information distribution has called attention to numerous archived collections in various institutions.

The University of Maryland Health Sciences Human Services Library (HS/HSL) launched, in May of 2011, the UM Digital Archive. Its purpose is to collect, preserve and distribute in digital format the academic works of the University of Maryland, the Founding Campus.¹

The Head of the UM Digital Archive obtained an award from the National Network of Libraries of Medicine, Southeastern/Atlantic Region for the digitization of the early dissertations for obtaining the degree of Doctor of Medicine at the University of Maryland.²

The goal for the digitization of these materials was twofold: discoverability and preservation. By providing remote access, people doing genealogical, medical, illustrative, or historical research, would have unlimited access to the collection. Preserving these documents digitally would allow enhanced access to the content without the possibility of damaging the original manuscripts.

This article discusses the historical context of the university and library, the initial process for preparing the dissertations for digitization and its subsequent revision, the unforeseen

challenges in metadata management and scanning quality control, and the rewards this project provided.

Historical Framework

The General Assembly of Maryland issued in 1807 *An Act for Founding a Medical College in the City or Precincts of Baltimore for the Instruction of Students in the Different Branches of Medicine*.³ As a result, the Maryland College of Medicine was established. The first students graduated in 1810. In 1812, with the addition of other faculties including Law, Divinity, and Art, the Maryland College of Medicine was re-chartered as the University of Maryland,⁴ thus becoming, in 1813, the founding institution of the present University System of Maryland.⁵

Part of the requirements for obtaining a degree of Doctor of Medicine from the University of Maryland during the nineteenth century was to submit a thesis in the Latin or English language, and students were to “publicly defend the same on the day of Commencement.”³ The dissertations from 1813 to 1889 have topics that reflect the medical interests and concerns of the time, such as Yellow Fever, Scarlatina, and Cynanche Trachealis (also known as Croup). Some include interesting calligraphy or hand drawn illustrations of anatomical parts, diseases, or medical procedures in pencil or color ink.

The early manuscript dissertations are an important part of the Historical Collections of the Health Sciences and Human Services Library, founded in 1813 with the purchase of Dr.

John Crawford's collection. Crawford was one of the members of a committee of physicians, including John B. Davidge, Nathaniel Potter, George Brown and others, appointed to plan and propose the creation of a medical college in Baltimore in 1803.⁴ Located in Baltimore City, the HS/HSL today serves the Schools of Medicine, Nursing, Pharmacy, Dentistry, Social Work and the Graduate School.

In 1938, Ruth Lee Briscoe, then Head Librarian, mentions in the *School of Medicine Bulletin* that the theses were bound in 189 volumes.⁶ These were rebound during the early 1940's. The HS/HSL Historical Collections still hold a few examples of dissertations bound individually, some by hand, and in several cases adorned with ribbon. The vast majority of these dissertations are now bound together in 162 volumes, in groups of 15 to 43. A table of contents (TOC), in typescript, presumably compiled around the time of the re-binding, lists the authors and titles of the original manuscripts for that volume. The header on the TOC page reads "University of Maryland Theses..." followed by a date or dates. The same date(s) appears on the spine of the volume. The theses are bound without regard to alphabetization of author, title, or subject matter. This irregular arrangement of the dissertations made the digitization project much more involved than planned.

Process

Providing metadata at the manuscript level for the volumes is an important element for discoverability. Most users look for a specific author, title, topic, or illustration. The original plan for assigning metadata was a simple 6-step process:

1. Populate the description field within a metadata spreadsheet provided by the vendor with information on authors and titles taken from the TOC.
2. Name each volume as it appeared in the TOC: "University of Maryland Theses..." followed by the year given on the TOC and the spine of the volume.
3. Confirm the Library of Congress Authority records for the University of Maryland School of Medicine for the time period in question.
4. Indicate the year. The year field would include the same year as the title, unless there were multiple years in one volume, then, the earliest date would be chosen.
5. Clarify author names. Some of the theses were written in Latin, as was common at the time, and the authors' names were Latinized as well. Another common practice was to write given names in an abbreviated form, for example "Geo." for George. If the purpose was for these authors to be discoverable, it was important to provide their names in the vernacular, and to spell out the complete name.
6. Identify the title. For titles that appeared in the TOC simply as "Inaugural Dissertation," it was necessary to go to the original manuscript to find the complete title, or to decipher the subject to add to the general title, for example, "Inaugural Dissertation (on Dysentery)."

The metadata was entered into a spreadsheet to be sent with the physical volumes of manuscripts to the scanning vendor.⁷ This spreadsheet, adapted for batch loading to the DSpace platform, would also be used to upload the metadata to the UM Digital Archive.

A seven member team dedicated to the project was recruited from the Resources and the Services Divisions of the library. Under the leadership of the Metadata Librarian, the team was trained to assign metadata in the spreadsheet format. A work schedule for the participants and a time frame for shipping the volumes was established. After a few weeks, the first shipment was ready to be sent.

Metadata Challenges

A problem was discovered while working on a volume, “University of Maryland Theses 1847,” for the second shipment. This volume included a case where the title appeared on the TOC as “Inaugural Dissertation.” The title page calligraphy was so complicated that it seemed impossible to decipher the title of the thesis. At this point the *University of Maryland Medical Faculty Matriculation List*, a handwritten ledger that contains the names of the faculty, matriculated students, and candidates for graduation for each year, from 1821 to 1892, was consulted.⁸ For most years, the titles of the dissertations of the candidates for graduation appeared next to their names. The author of the dissertation in question, James E. Porter Boulden, did not appear in the 1847 list.

Another source, *University of Maryland 1807-1907* by Eugene Cordell⁹, was turned to. It was discovered that the author had not graduated in 1847, but in fact three years later, in 1850. This year was validated by a third source, the *Catalogue of the Alumni of the School of Medicine*.¹⁰ Once corroborated, the title was located in the “Matriculation List” for the year 1850, then compared to the decorative calligraphy on the original title page shown in

image [1]. The triangulation process was successful, as the title of Boulden's thesis was finally clear: "Revulsion."

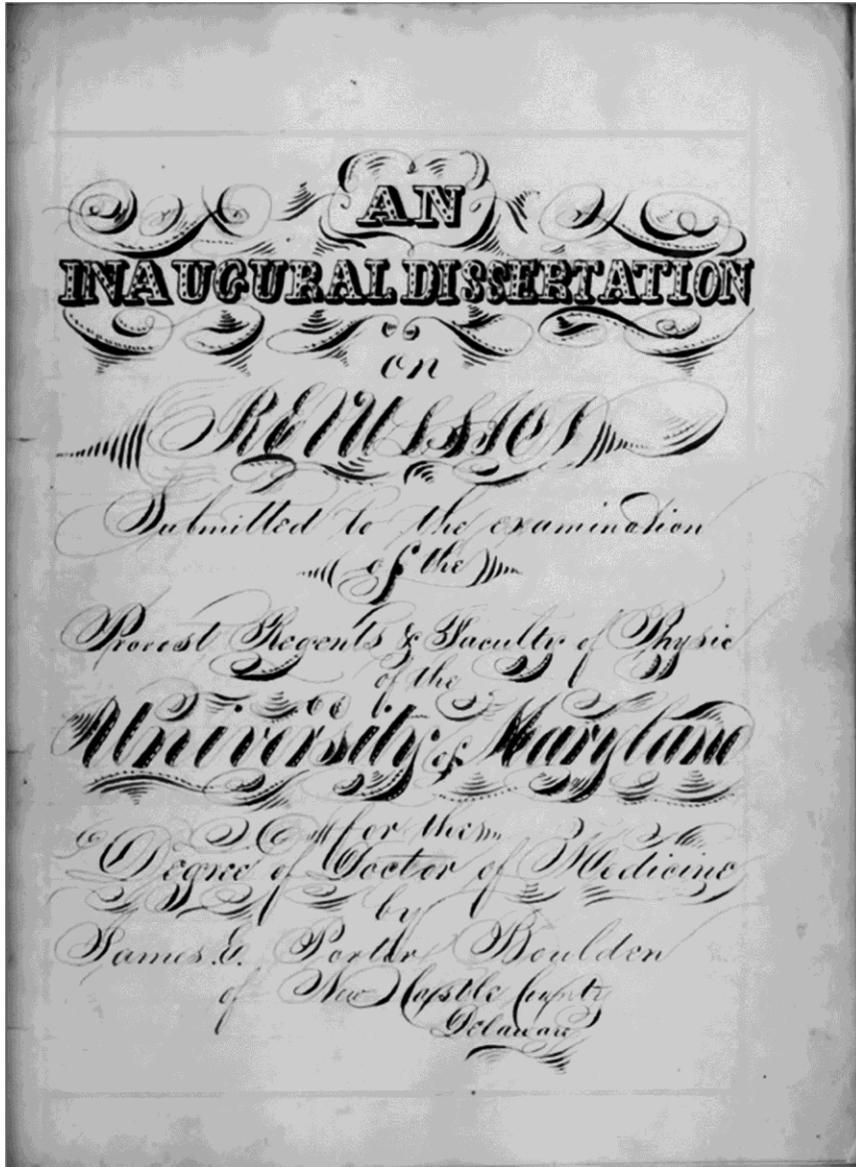


Figure 1: Boulden, J. An Inaugural Dissertation on Revulsion (1850). Bound in: University of Maryland Theses 1847, 1850(a). Grayscale of digital version: Internet Archive <<http://archive.org/details/universityofmary50unse>> PDF p. 889.

The question arose: were the discrepancies found in this thesis a sole exception? The team decided to verify the accuracy of information for other dissertations. In less than an hour four more cases of incorrect dates, plus errors in the transcription of the authors' names,

were found. It became clear that it was not enough to rely on the TOC to provide the correct metadata. These findings led to a decision to halt the first shipment that was ready to be sent for scanning. The process needed to be revised.

Process Revisions

The revised procedures included steps to ensure accuracy:

- 1) The information on the TOC for each volume was to be checked against the following sources:
 - a. Original Dissertation
 - b. University of Maryland Medical Faculty, Matriculation List 1821-1851; and 1851-1892.
 - c. Cordell, Eugene F. University of Maryland, 1807-1907. New York: The Lewis Publishing Company, 1907, Volume 2.
 - d. Catalogue of the Alumni of School of Medicine. Baltimore, Md.: Kelly, Piet & Co., 1877.
- 2) The inaccuracies were corrected in the metadata spreadsheet and the TOCs. Since the tables of contents were not part of the original manuscripts, they were altered in pencil. Many corrections needed to be made on the TOC of every volume examined. Since multiple pencil corrections could become confusing, a decision was made to create a new corrected TOC for every volume. These new tables of contents would allow for optical character recognition (OCR) increasing discoverability, as the

manuscripts are not machine-searchable. Each new TOC would be tipped into the volume. The legacy TOC would remain as well.

- 3) In order to document the references consulted, as well as the decisions made to correct the TOC, a page explaining the project was inserted at the beginning of each volume. Images [2] and [3] show the project page and an example of a new TOC referencing the resources used.
- 4) If a volume included theses of years other than those appearing on the spine and TOC heading, those years would be added next to the author's name in the TOC. The title on the TOC and the metadata spreadsheet would be similarly edited to include the correct years.

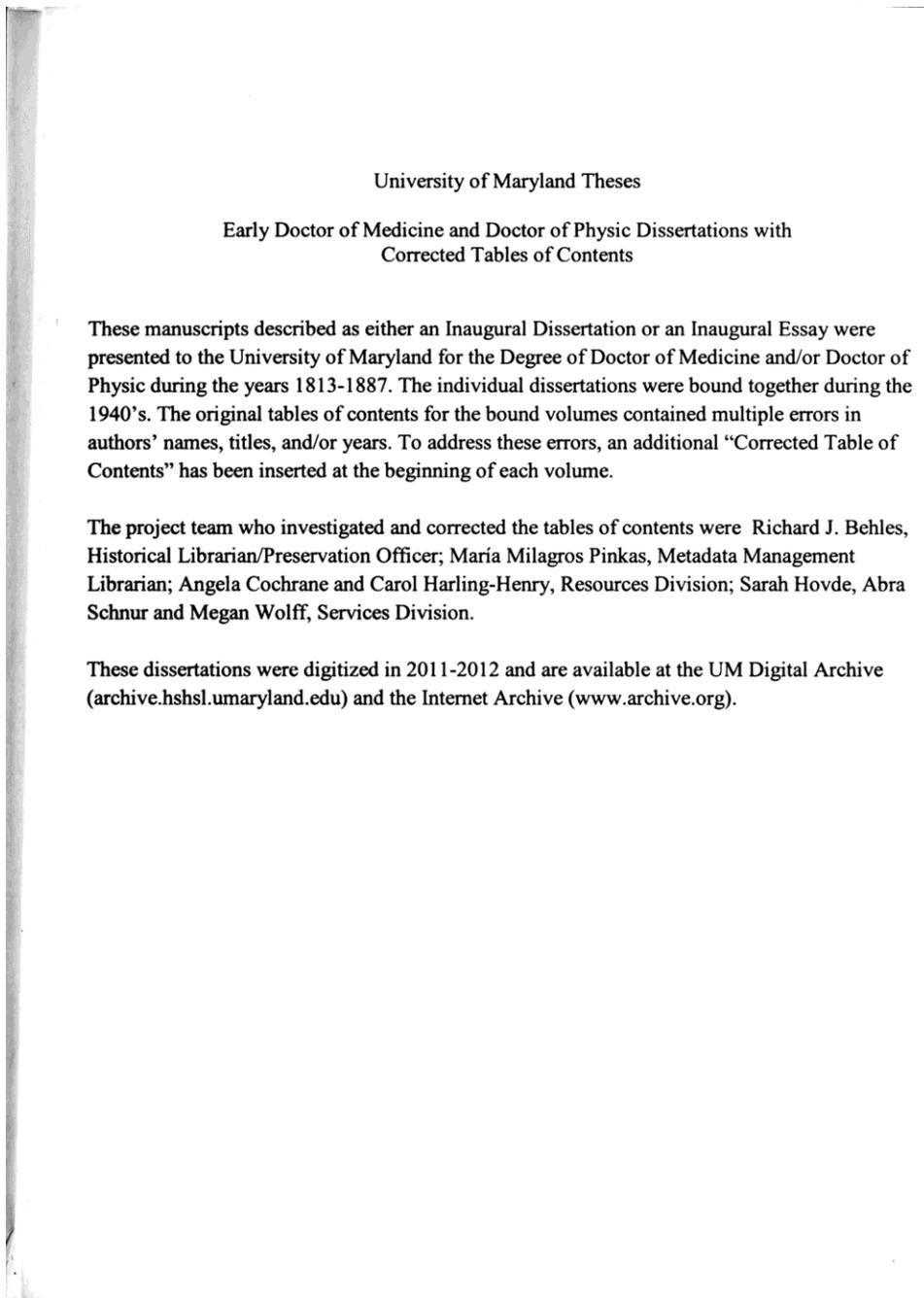


Figure 2: Example of project page inserted in each volume. University of Maryland Theses 1847, 1850 <<http://archive.org/stream/universityofmary50unse#page/n5/mode/2up>> PDF p.7

(CORRECTED TABLE OF CONTENTS)
 UNIVERSITY OF MARYLAND
 THESES
 1847, 1850 (a)

Author	Title	Year
Revell, William	Typhus Fever	1847
Mitchell, Lemuel P.	Gastritis, Acute and Chronic	1847
Kinkle, J.C.	Habit of Constipation	1847
Mathews, Alexander	Pneumonia	1847
Chapman, John S.	Inflammation	1847
Dougherty, Bernard A.	Pneumonia	1847
Mackall, Richard	Cold	1847
Pierce, W. Allen	Tetanus	1847
Smith, Thomas H.L.	Dysentery	1847
Laroque, Alfred	Chlorosis	1847
Bosley, Grafton M.	Sequalae of Malarious Fevers	1847
Marmillion, Edmond B.	Pneumonia	1847
Reindollar, William	Abortion	1847
Marcy, Virgil M.D.	Intermittent Fever	1847
Doyle, John A.	Neuralgia	1847
Boulden, James E. Porter	Revulsion	1850
Gilbert, George M.	Opium	1847
Dyson, Robert	Chemistry	1847

HSLSL 2011 for the UM Digital Archive. Sources consulted for corrections: Original Dissertation; University of Maryland Medical Faculty, Matriculation List, 1821-1851; Cordell, Eugene F. "University of Maryland, 1807-1907" (New York : The Lewis Publishing Company, 1907), Volume 2.

Figure 3: Example of new table of contents with referenced sources. University of Maryland Theses 1847, 1850 <<http://archive.org/stream/universityofmary50unse#page/n7/mode/2up>> PDF p.9

While using this more cautious procedure, a new problem was discovered: many of the volumes had serious binding errors, ranging from cut-off margins and folded pages to disordered theses. In some volumes, this was a matter of only a few pages bound out of

order, and was easily discovered and noted in the metadata; in others, sections from multiple theses were intermixed with no warning; and in some cases, different pages of the same thesis were erroneously bound in two different volumes, inserted in between the pages of another thesis. This new problem required a rethinking of procedures once again, to incorporate a step to decipher handwriting/subject matter to match pages bound out of order. The binding irregularities were evidenced in the new TOC's referencing the user to the location of the pages bound out of order. The process was no longer limited to transcribing metadata, but had become an intense research effort.

Scanning Quality Challenges

Once the first shipment was available in the Internet Archive and the UM Digital Archive, the project was featured in the HS/HSL newsletter, *Connective Issues*.¹¹ The metadata team searched for interesting illustrations and title pages to include in the column. This search led to the unexpected discovery of problems with the PDF versions of the theses.

The Internet Archive has several options for viewing digital items including "Read Online", "PDF", and e-book readers. The Read Online version was a satisfactory reproduction of the originals, the lines were crisp and detailed, and the contrast adequate. However, many of the PDF versions were blurred or faded to the point of illegibility, and all seemed to lack the necessary contrast and clearness of lines to render a faithful representation of the original. This had not been the case in a test volume sent for scanning at an earlier date, where the PDFs were clear. The physical items for this first shipment had been returned from the

scanning center, so the discrepancy between the PDFs and the original was confirmed. The digitization of the second shipment, already in the scanning center, was stopped until the PDFs of the first shipment were brought up to standards.

Proving the Case for Quality PDFs

The quality of the PDF versions is important to prospective users. The digitized manuscript volumes are not books meant to be read from beginning to end. Users are likely to be interested in one specific manuscript, author or illustration, not in the entire volume.

After a long process of identifying scanning errors throughout the 16 volumes in the first shipment, two of the volumes with multiple scanning errors were sent back to the scanning center. An illustration in one of the volumes became a perfect example for quality control checking. As shown in image [4], it was hand drawn in pencil and included the initial of the first name and the last name of the author. Parts of the drawing were only one line, and when that line was not visible, the drawing made no sense. Also, when the quality was poor, the author's signature was not visible.

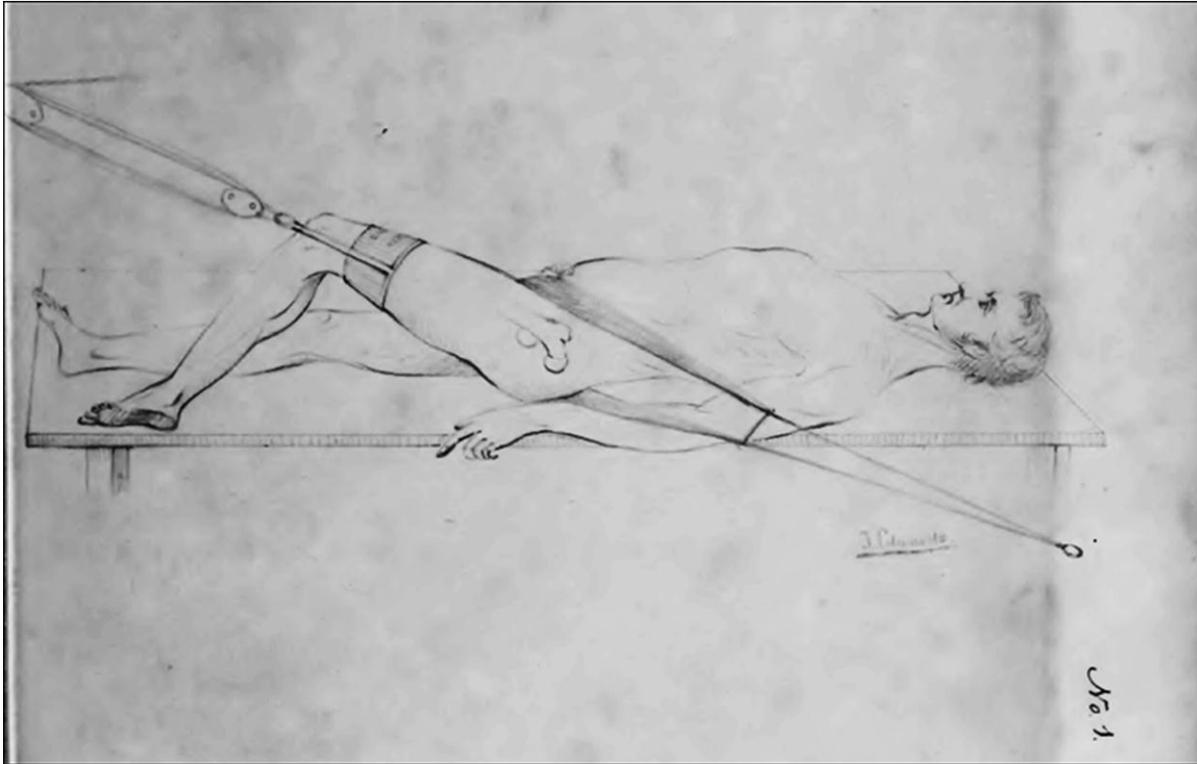


Figure 4: Edwards, T. A contrast between Smith and Cooper's mode of reducing dislocated Femurs (1831) Bound in: University of Maryland Theses 1831(a). Grayscale of digital version: <http://archive.org/details/universityofmary31unse> PDF p. 244. Accessible also from the UM Digital Archive <http://hdl.handle.net/10713/742>.

These two volumes were re-scanned by the scanning center with new parameters. The resulting PDFs were checked against the errors that had been identified, and found to be satisfactory. For the remaining volumes, the vendor proposed a post-scanning solution of adjusting the lighting. This option proved unsuccessful. A second quality control venture was started, in order to provide evidence that re-scanning was necessary for all the volumes in the shipment. These theses are one-of-a-kind manuscripts with historical value. A reliable reproduction of the original was extremely important.

Conway's "Table 1: Error model for digitized books"¹² served as a reference tool and provided the language and criteria to name the errors found. A checklist of common errors was created so that there was consistency in the evaluation.

Scanning Error Checklist:

- Portions not visible (cleaned)
- Crop (gutter)
- Warp (skewed)
- Illustration (tone, brightness, contrast)

Paper checklists were inserted as a marker for the physical pages that were not faithfully represented in the PDF version. A spreadsheet was created to compile the errors to share with the vendor, and to act as a reference for performing quality control of the PDFs.

After examining the data provided, the vendor agreed that the remaining volumes in the shipment needed to be re-scanned with the same scanning parameters. These parameters were re-adjusted until common factors that rendered the theses illegible were corrected, including primarily the cleaning, contrast, and gutter issues. The new adjustments were applied to the theses in the second shipment, and the first shipment was resent for re-scanning with these parameters as well.

Conclusion

From the initial process to the later challenges and adaptations, the digitization project provided learning experiences for staff and benefits for users. Staff became familiar with

reference tools and processes that assisted with assigning accurate metadata for authors, titles and dates for each thesis. Quality metadata and new tables of contents that can be OCRed provide users with more options for discovery.

The process of identifying and documenting scanning errors resulted in an applied lesson for staff on PDF scanning issues. The result was improved PDF versions, providing users with alternate viewing options.

The University of Maryland Early Dissertations for Doctor of Medicine is no longer a hidden collection.¹³ All 3,035 theses from 1813 to 1889 are accessible from any web browser, via the UM Digital Archive or the Internet Archive. The project, with all of its setbacks, evolved into a learning experience that will benefit future digitization projects of the HS/HSL.

References

¹University of Maryland, the Founding Campus. Digital Archive. "What is the UM Digital Archive?" Available: < <http://archive.hshsl.umaryland.edu> >. Accessed: September 13, 2012.

²This project has been funded in whole or in part with Federal funds from the National Library of Medicine, National Institutes of Health, Department of Health and Human Services, under Contract No. HHS-N-276-2011-00004-C with the University of Maryland, Baltimore, Maryland.

³*An Act for Founding a Medical College in the City or Precincts of Baltimore for the Instruction of Students in the Different Branches of Medicine: Passed November Session, 1807.* Baltimore, Md.: Pechin, 1807. Available: < <http://hdl.handle.net/10713/21> >. Accessed: July 13, 2012.

⁴Woodward, Theodore E. *Department of Internal Medicine, University of Maryland School of Medicine, 1807-1981.* Baltimore, Md.: Medical Alumni Association of the University of Maryland, Inc., 1987. Available: < <http://archive.org/details/departmentofinte00wood> >. Accessed July 13, 2012.

⁵University of Maryland School of Medicine. "Our History." Available: < <http://medschool.umaryland.edu/history.asp> >. Accessed: September 14, 2012.

⁶Briscoe, Ruth Lee. "Library of the School of Medicine, University of Maryland." *Bulletin of the School of Medicine* 22, no. 4 (April 1938): 172. Available: <<http://archive.org/stream/bulletinofuniver22unse#page/n151/mode/2up>>. Accessed: July 13, 2012.

⁷The Health Sciences and Human Services Library is a member of the Lyrasis Mass Digitization Collaborative < <http://www.lyrasis.org/Products-and-Services/Digital-Services/Mass-Digitization-Collaborative.aspx> >. It is through this Lyrasis partnership that the digitization by the Internet Archive's scanning center is coordinated. Accessed: September 14, 2012.

⁸University of Maryland Medical Faculty, Matriculation List 1821-1851; and 1851-1892.

⁹ Cordell, Eugene F. *University of Maryland, 1807-1907.* New York: The Lewis Publishing Company, 1907, Volume 2.

¹⁰*Catalogue of the Alumni of School of Medicine.* Baltimore, Md: Kelly, Piet & Co., 1877.

¹¹“Digital Archive Highlight.” Connective Issues: The E-Newsletter of the University of Maryland Health Sciences and Human Services Library 6, no. 3 (January/February 2012).

Published online. Available:

<<http://www.hshsl.umaryland.edu/general/news/newsletter/index.php/2012/02/#13>>.

Accessed: September 14, 2012.

¹²Conway, Paul. “Archival Quality and Long-Term Preservation: A Research Framework for Validating the Usefulness of Digital Surrogates.” *Archival Science: International Journal on Recorded Information*. Published online: (September 21, 2011) doi: 10.1007/s10502-011-9155-0 . Available:

<<http://www.springerlink.com/content/9322j77m6327h123/fulltext.html>>. Accessed:

September 6, 2012.

¹³The handle to the School of Medicine – “Early Dissertations of the University of Maryland Collection” in the University of Maryland Digital Archive is

<<http://archive.hshsl.umaryland.edu/handle/10713/687>>. Accessed: [September 10, 2012](#).

ABOUT THE AUTHORS

María M. Pinkas, M.L.S. (mpinkas@hshsl.umaryland.edu) is Metadata Management Librarian in the Resources Division; Abra Schnur, B.A. (aschnur@hshsl.umaryland.edu) is Reserves Technician and Megan Wolff, B.A. (mwolff@hshsl.umaryland.edu) is Reserves Associate in the Services Division; Sarah Hovde, B.A. (shovde@hshsl.umaryland.edu) and Carol Harling-Henry (charling@hshl.umaryland.edu) are Library Technicians in the Services and Resources Divisions, respectively, of the Health Sciences and Human Services Library of the University of Maryland, Baltimore, Maryland, USA