Managing Electronic Theses and Dissertations in Acquisitions

Michelle Flinchbaugh

INTRODUCTION

In February 2009, the acquisitions librarian for the University of Maryland, Baltimore County (UMBC), proposed that the Acquisitions Unit take responsibility for digital transfer services for digital collections. The workload of the Acquisitions Unit had been declining, and all indications were that it would continue to decline. Ordering processes had been streamlined and required less staff time, and purchasing of e-book packages and a switch to patron-driven acquisition for e-books suggested that staff time spent on ordering would continue to diminish. An article about transfer-related services suggested a new set of services the unit could provide:

- acquiring digital content;
- conducting quality review of digital content;
- moving digital content between systems; and
- inventorying, manipulating, and ingesting digital content into digital collections.[1]

UMBC is a public research university that enrolls approximately 13,500 students. Founded in 1966, it's a very young but well-rated university, one of 147 U.S. universities named as a top global university by *U.S. News and World Report*; it has also received national recognition as a top university in innovation and teaching. Approximately

2,500 of UMBC's students are grad students. UMBC's library, the Albin O. Kuhn Library and Gallery, uses CONTENTdm to house its digital collection. A primary emphasis of UMBC's Special Collections is photography and the history of photography, and CONTENTdm was originally purchased to house digitized photographs. However, over time the collections on CONTENTdm have grown to include records, proceedings, reports, digitized books, and electronic theses and dissertations (ETDs). Some of the material was digitized from print collections, but a good portion of the documents are submitted in digital format, including some ongoing serial publications. It was these latter materials that the acquisitions librarian was suggesting that Acquisitions manage.

A more pressing need was the management of ETDs. All UMBC theses and dissertations are submitted to ProQuest, and the submission process is managed by UMBC's Graduate School. ProQuest publishes the ETDs and provides them to the library in electronic format, along with metadata in XML format. The work formerly belonged to the University Archives in the Special Collections Department but had been passed to a serials librarian, and scripts had been developed to reformat the XML metadata for CONTENTdm. The librarian managing the process had resigned and the scripts had broken. They were moved to the archivist in Special Collections. Moving the XML metadata into CONTENTdm was a challenge with the broken script. The archivist had a student manually copy and paste metadata elements from the XML files into CONTENTdm. The process of locating a specific XML tag to copy its contents into a particular CONTENTdm field was time-consuming.

The acquisitions librarian had some experience in managing and manipulating large data sets. She believed she could probably automate the process using Microsoft Access or Excel features and then train her staff to carry them out. She leveraged an article, "Repurposing ProQuest Metadata for Batch Ingesting ETDs Into an Institutional Repository," about how another library had done this, which gave her some ideas about how it could be achieved.[2] The acquisitions librarian, archivist, and head of Technical Services all met, information was provided, and the acquisitions librarian took over managing the ETDs.

TAKING RESPONSIBILITY FOR THE ETDS IN ACQUISITIONS

The acquisitions librarian initially attempted to use Access for the metadata conversion process but soon realized this would not work because of Access's limits on field size. She discovered that Excel has XML utilities but that they are included in a Developer tab that is turned off by default upon installation. Using the XML utilities, formulas, and macros, she could develop an Excel template for changing ProQuest XML metadata into a format that could be ingested by CONTENTdm.

After a minimal period of testing, the ongoing processing was then handed off to an acquisitions technician. The technician, who is computer savvy, learned a number of new software packages, as well as new Excel features, for handling the ETDs, including FileZilla for downloading files, 7-Zip for unzipping files, Adobe Acrobat for manipulating pdfs, and the CONTENTdm Client for loading files onto the CONTENTdm server. While she was quite able to handle this complex procedure, many exceptions were discovered. Often an ETD could not be processed using the general procedure, so the acquisitions librarian had to investigate and expand the procedure to cover special cases. After a period of about a year, nearly all exceptions were documented—although the technician occasionally still discovers new variations.

Soon after taking over work on the ETDs, it became apparent that managing ETDs would involve more than just devising a system for ingesting metadata into CONTENTdm. The following additional tasks also materialized:

- The library wanted to enter into negotiations with the Graduate School to eliminate personal information from ETDs.
- The library wanted to enter into negotiations with the Graduate School to develop a mechanism to provide information about what theses and dissertations were expected and when.
- The library wanted to enter into negotiations with the Graduate School to create and implement a permissions form as part of the ETD submission process that would permit them to be publicly accessible.

- Receiving and paying for print theses and dissertations needed to be moved from the Graduate School to the library.
- Procedures needed to be developed for the occasional older theses and dissertations Special Collections received permission to digitize and make accessible.
- Inconsistencies between the ETD collection and other CONTENTdm collections had to be investigated and resolved.
- A plan for maintaining the integrity of URLs in the catalog had to be put into place.
- A method for loading ETD metadata into OCLC catalog records had to be devised.
- Upon the implementation of an institutional repository, the ETDs had to be migrated from CONTENTdm to DSpace.

Collaboration would be required with the archivist in Special Collections, the Graduate School, and a cataloger. Further, information would have to be obtained from ProQuest to better understand the metadata, particularly regarding embargo information. Some issues also impacted the Interlibrary Loan Department, as it occasionally lends theses and dissertations with permission from the authors.

LITERATURE REVIEW

The earliest work on ETDs occurred in the late 1980s, with meetings attended by representatives of University Microfilms International (UMI), the Coalition for Networked Information, the Council of Graduate Schools, Virginia Tech, and the University of Michigan. In the early 1990s the same group started a project to develop standards and applications for capturing ETDs electronically. An outcome of that project was UMI's ProQuest platform for digital dissertations that began accepting all submissions in electronic format in 1997. UMI also began scanning and digitizing paper and microform submissions. The Southeastern Universities Research Association (SURA) provided funding for Cornell University, the University of Michigan, Penn State, and Virginia Tech to develop and disseminate a standard method of using SGML to make dissertations available online.[3]

The U.S. Department of Education funded a grant for Virginia Tech to create a national digital library of ETDs, which became the Networked Digital Library of Theses and Dissertations (NDLTD). The NDLTD is a confederation of member institutions and organizations that provides access to theses and dissertations. With additional funding from Adobe, IBM, and Microsoft, Virginia Tech became the worldwide leader in ETD development.[4] In 1999, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) became interested in ETDs. In 2000, UNESCO supported Virginia Tech in developing an ETD best practices document, *The Guide for Electronic Theses and Dissertations*.[5]

Professional organizations in both higher education and librarianship began publishing informational articles on ETDs. In 2003, Hall, Hoover, and Wolverton authored a work published in *Technical Services Quarterly* covering the history of ETD initiatives and recommending campus planning and networking with Virginia Tech to institutions interested in developing an ETD initiative.[6] Joan K. Lippincott, with the Coalition for Networked Information, wrote *Institutional Strategies and Policies for Electronic Theses and Dissertations*, published in the 2006 Educause Center for Applied Research *Research Bulletin*.[7] Fyffe and Welburn wrote "ETDs, Scholarly Communication, and Campus Collaboration: Opportunities for Libraries," published in *College & Research Libraries News* in 2009.[8]

Articles also appeared focusing on the development of ETDs in different regions. "The Development and Promotion of Electronic Theses and Dissertations (ETDS) Within the UK" by Copeland and Penman, published in 2002, provides information on three projects funded by the Joint Information Systems Committee to develop ETDs in the United Kingdom, focusing, in particular, on the project led by Robert Gordon University.[9] "University of Waterloo Electronic Theses: Issues and Partnerships," appearing in 2006 in *Library Hi Tech*, describes a project untaken jointly by the University of Waterloo, Theses Canada, and NDLTD to provide Open Access ETDs in Canada.[10] "Electronic Theses and Dissertation (ETD) Repositories: What Are They? Where Do They Come From? How Do They Work?," written by Kristin Yiotis and published by *OCLC Systems & Services* in 2009, covered their development in the United States and included recommendations for libraries wishing to implement them as they proliferated and as different models and systems emerged.[11]

Some authors focused on detailed accountings of their institution's ETD implementation. Park, Zou, and McKnight wrote about McGill University's pilot ETD submission project; their article was published in 2007 in the journal *Program: Electronic Library and Information Systems*.[12] Bishop, Marshall, and Winter published an article about the University of Central Florida's ETD implementation in the Educause Center for Applied Research *Newsletter* in 2007.[13] Wang, Bulick, and Muyumba wrote an article about initiating an ETD program at Indiana State University that was published in 2014 in *OCLC Systems & Services: International Digital Library Perspectives*.[14]

Hall, Hoover, and Wolverton surveyed U.S. ETD programs to analyze trends and review models. They found that in most cases, ETDs where managed collaboratively by the graduate school and the library, most often with the graduate school taking the lead in coordinating activities. The involvement of an array of other units was also mentioned. They also surveyed the position title of the individual coordinating ETDs, the number of employees working on them, the software used, and more.[15] Early and Taber surveyed North Carolina libraries with ETD programs, investigating how they handle collaboration, workflows, and division of labor in ETD programs. They found that ETD programs require a large variety of skill sets and commonly involve multiple departments, usually both libraries and graduate schools, and that collaboration is usually necessary.[16]

Many articles have been published on cataloging ETDs, but coverage of these is outside of the scope of this article. Instead, we focus on reports of methods of manipulating metadata. McCutcheon, Kreyche, and Maurer published an article about moving ETDs from the centralized OhioLINK ETD Center to libraries' catalogs in *Library Hi Tech* in 2008. They used OAI–PMH (Open Archives Initiative–Metadata Harvesting Protocol) to extract metadata and Perl programming to modify and enhance the data, converting it from ETD-MS to MARC 21 and inserting it into their Innovative Interfaces catalog.[17] Averkamp and Less wrote about their process for batch ingesting ETDs into an intuitional repository utilizing XSLT, published in *Code4lib Journal* in 2009.[18] Amanda Z. Xu talked about a similar process at the 2016 ALCTS (Association for Library Collections and Technical Services) Technical Services Workflow meeting at the Midwinter American Library

Association Meeting.[19] At the same meeting, Steven H. Holloway presented on a process to convert Excel ETD data into MARC Bib and NACO records.[20]

WORKING WITH THE GRADUATE SCHOOL

Early on, the acquisitions librarian set up a meeting with appropriate people from the Graduate School and the head of Technical Services. The library's primary issues were removing personal information from the ETDs, knowing what the library should receive and when, and getting permissions from authors to make their documents publicly accessible. These issues were discussed at the meeting but were not resolved. The Graduate School, in turn, wanted the library to begin paying for the print theses and dissertations that it had been purchasing for Special Collections. This was arranged, and Acquisitions also took over this work, including a backlog of print theses and dissertations that the Graduate School had received but had not paid for. The library's issues would be raised again with the Graduate School periodically over the years, but Graduate School staff generally were too busy to meet with the library again.

The acquisitions librarian also became the point person for students with questions about the publication of theses and dissertations, which both Special Collections and the Graduate School began forwarding to her. While not a substantive workload, these questions proved to be a helpful opening in getting at least one of the library's issues addressed. A recent graduate who was job hunting had a URL to his thesis in his resume. At an interview, a prospective employer told him they wanted to see it but the link didn't work. The acquisitions librarian helped the student by immediately having him sign a permissions form and making his document publicly accessible. In addition, she renewed her request for the Graduate School to put a permission document in place so that the theses and dissertations could be made publicly accessible. The Graduate School finally agreed to this, cautioning that it would have to take the issue to its directors and the Graduate Student Association before implementing.

To design the actual permissions form, the acquisitions librarian obtained a copy of the permissions form the University of Maryland,

College Park, was using from its repository manager, and this was passed back and forth between the Graduate School and acquisitions librarian in rounds of editing. The permissions document constituted an Open Access mandate for the ETDs—embargoes were allowed, but there was no opt-out. The acquisitions librarian was asked to attend the Graduate School directors meeting, where the issue would be discussed. She did a short presentation and the form with the Open Access mandate passed unanimously. The final version of the form was sent to campus counsel for a legal review and was approved. When many months had passed without hearing anything further, the acquisitions librarian inquired about where this stood with the Graduate School. A new version of the permissions form with an opt-out addition was provided, so the Open Access ETD mandate was not going to happen at UMBC; however, the students would at least have the opportunity to choose to make their thesis or dissertation Open Access.

Many more months passed without a word and without receiving any permissions forms. The Graduate School said it thought it had been set it up so that ProQuest would get the form completed as part of the submission processes. Several years later, the acquisitions librarian began receiving paper permissions forms in campus mail from the Graduate School. To date, those remain unprocessed with no procedure in place, as they arrived just after migrating the ETDs to a new platform and immediately thereafter the author went on research leave.

In 2015, UMBC's library director retired. The hire of a new library director and his meeting with a new director of the Graduate School proved to be another opening to address more of the library's issues with the Graduate School, and a mechanism was finally put in place to inform the library of what it should be receiving and when.

WORKING WITH LIBRARY COLLEAGUES

During the long periods of waiting for the Graduate School, the acquisitions librarian was working on issues that didn't require library–Graduate School collaboration.

As soon as the decision had been officially made to begin requiring students to complete a permissions document, a plan was developed

for transitioning the collection from one in which no items were publicly accessible to one in which some were accessible and some weren't. The existing ETDs had to have access restrictions put on them individually, along with a note indicating that the item isn't publicly available but could be provided via interlibrary loan with the author's consent. Special Collections agreed to have a student assistant put the access restrictions and note on the existing ETDs, and the acquisitions librarian provided step-by-step instructions. Once that was completed, the collection was made publicly accessible.

Along the way, Special Collections wanted to occasionally obtain permission from authors of older ETDs to digitize their documents and make them publicly accessible when frequently requested via interlibrary loan. The archivist adapted the form developed for ongoing use by the Graduate School for this purpose and began occasionally sending the acquisitions librarian older digitized dissertations (for which they had obtained permission to make them publicly accessible) for inclusion in the ETD Collection in CONTENTdm. Eventually, they agreed to place all completed license forms on a shared drive directory with access limited to interested parties—to include Acquisitions, Interlibrary Loan, and Special Collections. Since the digitized documents do not come with metadata and are missing several key pieces of information that are normally included, the acquisitions librarian developed a separate procedure for them. She determined that the best method of obtaining metadata was to utilize the catalog record for the item. She developed a procedure based on copying and pasting key bits of information from the catalog record into the appropriate CONTENTdm fields. At present, Acquisitions receives less than of five of these per year, so the copy and paste procedure suffices, but if the volume substantively increases, an automated mechanism for handling these will need to be developed.

During this time, the acquisitions librarian also developed a plan for modifying the ETD metadata schema in CONTENTdm to better match the other collections. First, she checked with the archivist about the discrepancies in case there was a reason for them that she didn't understand. She investigated the CONTENTdm administration module's capacity for making global changes and wrote up a plan for changes that she knew the system would facilitate. She sent the plan via email to the Special Collections librarians, catalogers, and head of

Technical Services for feedback. After the feedback period had expired and she had received no objection to the changes, she went ahead and made them.

Also during the long period of waiting for the Graduate School, Cataloging began adding ETD URLs to the catalog records for the print copy. The acquisitions librarian, in consultation with the head of Technical Services, investigated some possible methods of moving the ETD records from CONTENTdm to OCLC, but none seemed suitable. Around the same time, the acquisitions librarian heard about XSLT at a conference, and she and one of the catalogers attended a class on XSLT. The other cataloger in the library quickly followed suit and also learned XSLT. The acquisitions librarian reprogrammed the ETD metadata to include XSLT. Once this was done, she began looking at the ability to utilize XSLT to reformat data to import into MarcEdit for batch modification of metadata to create bibliographic records for import into OCLC and the catalog. On testing, this worked. However, she didn't have time for this, she wasn't familiar with MarcEdit, and her cataloging knowledge was obsolete. She therefore passed the information regarding how this could be done to the head of Technical Services, and a cataloger began to work on it.

The acquisitions librarian and cataloger discussed and determined that Cataloging would get the XSL metadata from the FTP server to which ProQuest sends the ETDs. The cataloger established that cataloging the ETDs would require a different XSLT process, but the acquisitions librarian provided her with acquisitions procedures and coding as a starting place so that the cataloger didn't have to remake the wheel. Soon after, a process was in place for utilizing the XSLS for cataloging as well, saving substantive time since individual records no longer needed to be manually created.

In 2018, UMBC implemented an institutional repository as a part of MD-SOAR, a consortial repository on the DSpace platform. Early in the implementation process, the acquisitions librarian thought it would be good to move the ETDs to the new platform and called a meeting that included special collections librarians, technical services librarians, and many department heads about doing so. All agreed with the platform move and discussed things that would need to be done along with the migration, such as redirecting web page and cataloging links and redoing LibGuides about ETDs. After the meeting,

the acquisitions librarian created a step-by-step plan for the migration of the ETDs, which was shared with all. She, along with acquisitions staff, utilized the slow period during the next summer to do quality assurance on the existing ETD collection in CONTENTdm, finding and correcting issues.

The following summer, the ETD's metadata was extracted from CONTENTdm and the actual object files in the ETD collection were provided by OCLC for load into DSpace. Differences in the way CONTENTdm and DSpace handle objects with multiple files required reformatting them. Differences in the way CONTENTdm and DSpace handle special characters required finding and replacing them. While working on the metadata, department names were also standardized to match the department names in the new repository. Additionally, problems with metadata were identified and corrected using Excel tools while the metadata was in a spreadsheet format.

After the successful load of the ETDs to a test server, they were loaded to DSpace. Thereafter, the acquisitions librarian rewrote the XSLT and Excel macros utilized for the ETDs to accommodate metadata differences in the two collections and to standardize departments' names automatically as part of the process. Catalogers redirected the catalog record links to DSpace. The web librarian redirected ETD web links to DSpace, and reference librarians worked on redoing a LibGuide on ETDs. After allowing time to ensure that all is complete and there are no problems, a final notification will be sent to all stakeholders and the ETDs in CONTENTdm will be made inaccessible. After another period of time—again to ensure there are no problems—the ETD Collection will be deleted from CONTENTdm.

CONCLUSION

Placing ETD processing in Acquisitions isn't something that every library will do, but this made sense at UMBC both because the skills of staff meshed well with the work that needed to be done and workloads had decreased. The work involved quickly snowballed and has required Acquisitions to collaborate with several other departments in the library, as well as with the Graduate School. This has called for patience, creative use of technology, resourceful use of

opportunities to get the Graduate School on board with changes, a very slow decision-making process requiring consultation with numerous stakeholders, and a great deal of understanding of others' priorities and perspectives. For the Acquisitions Department at UMBC, managing the ETDs is a collaborative endeavor that is always engaging and challenging.

NOTES

This chapter expands and updates Michelle Flinchbaugh's portion of the article "Acquisitions and the Digital Repository" (Steven Douglas and Michelle Flinchbaugh, "Acquisitions and the Digital Repository," *Against the Grain* 23, no. 3 (June 2011): 61–62, accessed July 20, 2017, https://mdsoar.org/handle/11603/190). Flinchbaugh's portion of the article was on loading electronic theses and dissertations (ETDs) into digital collections in CONTENTdm and has been expanded with new content.

1. Justin Littman, "A Set of Transfer-Related Services," *D-Lib Magazine* 15, no. 1/2 (2009), accessed July 20, 2017, http://www.dlib.org/dlib/january09/littman/01littman.html.

2. Shawn Averkamp and Joanna Lee, "Repurposing ProQuest Metadata for Batch Ingesting ETDs Into an Institutional Repository," *Code4lib Journal*, no. 7 (2009), accessed July 6, 2017, http://journal.code4lib.org/articles/1647.

3. Kristin Yiotis, "Electronic Theses and Dissertation (ETD) Repositories: What Are They? Where Do They Come From? How Do They Work?" *OCLC Systems & Services: International Digital Library Perspectives* 24, no. 2 (2008): 102, https://doi.org/10.1108/10650750810875458.

4. Ibid., 103.

5. "ETD Guide." NDLTD, accessed August 24, 2017. http://etdguide.ndltd.org/.

6. Susan Hall, Lona Hoover, and Robert E. Wolverton Jr., "Publishing Electronic Theses and Dissertations: Reconfiguring Library Services," *Technical Services Quarterly* 21, no. 2 (2003): 63–70, accessed August 23, 2017, https://doi.org/10.1300/J124v21n02_05.

7. Joan K. Lippincott, "Institutional Strategies and Policies for Electronic Theses and Dissertations," Educause Center for Applied Research

Research Bulletin, no. 13 (2006), accessed May 10, 2019, https://library.educause.edu/resources/2006/6/institutional-strategies-and-policies-for-electronic-theses-and-dissertations.

8. Richard Fyffe and William C. Welburn, "ETDs, Scholarly Communication, and Campus Collaboration: Opportunities for Librarians," *College & Research Library News* 69, no. 3 (March 2008): 152–55, accessed August 22, 2017, http://crln.acrl.org/index.php/crlnews/article/view/7952/7952.

9. Susan Copeland and Andrew Penman, "The Development and Promotion of Electronic Theses and Dissertations (ETDs) within the UK," *The New Review of Information Networking* 10, no. 1 (2004): 19–32.

10. Christine Jewell, William Oldfield, and Sharon Reeves, "University of Waterloo Electronic Theses: Issues and Partnerships," *Library Hi Tech* 24, no. 2 (2006): 183–96, https://doi.org/10.1108/07378830610669565.

11. Kristin Yiotis, "Electronic Theses and Dissertations: What Are They?"

12. Eun G. Park, Qing Zou, and David McKnight, "Electronic Thesis Initiatives: Pilot Project of McGill University, Montreal," *Program: Electronic Library and Information Systems* 41, no. 1 (2007): 81–91, https://doi.org/10.1108/00330330710724917.

13. Patricia Bishop, Ruth Marshall, and Debra Winter, "A Robust Thesis and Dissertation Program at UCF," *Educause Center for Applied Research Research Bulletin*, no. 2 (2007), accessed August 22, 2017, https://library.educause.edu/~/media/files/library/2007/1/erb0703-pdf.pdf.

14. Xiaocan (Lucy) Wang, Natalie Bulick, and Valentine Muyumba, "Publishing Student Scholarship: Exploring the ETD Initiative at a Medium-sized Institution," *OCLC Systems & Services: International Digital Library Perspectives* 30, no. 4 (2014): 232–48, accessed July 6, 2017, https://doi.org/10.1108/OCLC-02-2014-0008.

15. Susan Hall, Lona Hoover, and Robert E. Wolverton Jr. "Administration of Electronic Theses/Dissertations Programs: A Survey of U. S. Institutions," *Technical Services Quarterly* 22, no. 3 (2005): 1–17, accessed August 23, 2017, https://doi.org/10.1300/J124v22n03_01.

16. Mary G. Early and Anne Marie Taber, "Evolving Collaboration: Electronic Thesis and Dissertation Workflows in North Carolina,"

Collaborative Librarianship 2, no. 1 (2010), accessed July 6, 2017, http://digitalcommons.du.edu/collaborativelibrarianship/vol2/iss1/3/.

17. Sevim McCutcheon, Michael Kreyche, Margaret Beecher Maurer, and Joshua Nickerson, "Morphing Metadata: Maximizing Access to Electronic Theses and Dissertations," *Library Hi Tech* 26, no. 1 (2008): 41–57, https://doi.org/10.1108/07378830810857799.

18. Avercamp and Lee, "Repurposing Proqust Metadata."

19. Moreno, Heylicken, "Report of the ALCTS Technical Services Workflow Efficiency Interest Group Meeting; American Library Association Midwinter Meeting; Boston, January 2016," *Technical Services Quarterly* 33, no. 4 (2016): 432, accessed July 6, 2017, https://doi.org/10.1080/07317131.2016.1203648.

20. Ibid., 434.