COMMUNICATIONS ON PRACTICE

Digitizing Pre-1978 Dissertations at Binghamton University Libraries

Erin E. Rushton, Caitlin Holton, and Jamey McDermott

In 2023, Binghamton University Libraries initiated a project to digitize its pre-1978 dissertations and make them available in its institutional repository. This Communication on Practice provides an overview of the key decisions made before embarking on the project, the workflow, and the challenges encountered. We drew upon the experiences and lessons learned from other institutions to guide our process, and we hope this paper will serve as a resource for those considering similar projects at their institutions.

In 2016, Binghamton University Libraries established the Digital Initiatives and Resource Discovery Department. Initially comprising a single librarian, the department has since expanded to seven employees who provide leadership, direction, and innovation in digital scholarship, digital collections, digitization, digital preservation, scholarly communications, and oral histories. Among its many responsibilities are the oversight and management of the digitization lab, institutional repository (bepress Digital Commons), and digital preservation system (Ex Libris Rosetta).

In the department's early days, when there were fewer active projects, students were occasionally tasked with digitizing pre-1978 dissertations and adding the dissertations to the institutional repository. This idea was inspired by Gail Clement and Melissa Levine's 2011 article, "Copyright and Publication Status of pre-1978 Dissertations: A Content Analysis Approach."[1] They suggested that many pre-1978 dissertations were likely in the public domain due to a lack of copyright notice and a potential failure to renew copyright. Given the minimal risk in digitizing pre-1978 dissertations and the need to provide meaningful work to keep the student workers occupied, the project was initiated on an ad-hoc basis.

As the department matured, a digital initiatives assistant was hired to manage the Digitization Lab's daily activities and supervise student employees. The digital initiatives assistant also supported the institutional repository. During this period, the lab's capabilities increased to better accommodate the digitization needs of the materials housed in our collections. One of the most critical additions was an InoTec Scamax 631, a high-volume feed scanner that improved the scope and efficiency of projects by reducing the time required to scan documents.

Around this time, we noticed many downloads from the dissertations that the students had digitized and deposited into the repository. For instance, a dissertation from 1970 had more than 600 downloads. Although this dissertation was a bit of an outlier, many other dissertations demonstrated

Erin E. Rushton (erushton@binghamton.edu) is the Head of Digital Initiatives and Resource Discovery at Binghamton University Libraries, https://orcid.org/0000-0001-5303-4958. **Caitlin Holton** (cholton@binghamton.edu) is the Digital Initiatives Assistant at Binghamton University Libraries, https://orcid.org/0000-0002-2596-4303. **Jamey McDermott** (jmcderm5@binghamton.edu) is a PhD Candidate in English at Binghamton University, https://orcid.org/0009-0006-0956-4445.

Library Resources & Technical Services | April 2025
https://doi.org/10.5860/lrts.69n1.8441

significant usage, some receiving thirty to forty downloads. The interest in the dissertations was not particularly surprising. We had come across articles in the library literature that suggested that digitized dissertations receive far more usage than traditional print because of their broader accessibility and visibility. For instance, Daniel S. Dotson's article on Ohio State University's Electronic Theses and Dissertations (ETDs) project highlighted the usage patterns and impact of ETDs. At the time of publication, Ohio State had made more than 50,000 dissertations available online, with more than 29,341,996 downloads.[2]

Interest in this collection motivated us to consider the feasibility of a dissertation digitization project. There were other compelling reasons to consider such a project. First, the project aligned with the libraries' recently created strategic plan, which promotes amplifying campus scholarship and transforming collections.

Second, as open access advocates, we encourage faculty and students to share their scholarly and creative work, including ETDs, in the institutional repository. This project would populate the repository and enhance alumni scholarship and research visibility, especially since there has not yet been a lot of interest from students in depositing their ETDs in the repository. Third, the library's print dissertation collection occupied a substantial space on the second floor of the main library. Although there were no immediate plans to renovate or repurpose the space, using it for a low-circulating and less visually appealing collection was not ideal. The digitization and deaccession of the dissertations could free up space that we could use for study space or more high-use collections.

Finally, we found reassurance in the experiences of other libraries that had embarked on similar projects and had lived to tell the tale.

Project Considerations

Although we had compelling motivations to proceed, we also faced considerations that warranted caution. Our key concern was whether we collectively possessed the time and resources required for a project of this magnitude, which would likely be laborious and time-consuming. We also needed to evaluate the project's priority relative to other ongoing and planned projects and ascertain the level of administrative support. Although the Digital Initiatives Department would primarily be responsible for managing the project and conducting the work, other departments, including Cataloging, Preservation, and Reader Services, would be impacted.

To ensure sufficient support and interest, we prepared a proposal describing the project's goals and objectives, issues to address, and a tentative workflow and timeline. Additionally, we included the success stories we had collected from other institutions. We scheduled a meeting with Cataloging, Preservation, and Reader Services to discuss the proposal and address any concerns. Ultimately, the project was well-received, and we obtained the green light to move forward.

Project Decisions

Before starting the project, we made policy- and workflow-related decisions.

Opt-in or Opt-out for Authors

Our exploratory review revealed that some institutions had invested considerable effort in locating contact information of ETD authors to offer the option to opt in or out.[3] We also found examples of libraries that only contacted alumni authors after digitizing and depositing their work. Despite varying approaches, adverse reactions from alumni were rare in the literature. This evidence led us to not seek permission from individual authors proactively. Instead, we included a link to the libraries' take-down policy on the collection landing page to facilitate the removal of a dissertation on request.

Scope of Project

Clement and Levine's 2011 article suggests that pre-1978 dissertations are likely in the public domain if the author did not renew the copyright.[4] Given this analysis, we focused the project's first phase on pre-1978 publications. In total, 293 dissertations met these criteria.

Retention of Print Copies

The proposed workflow involved unbinding the physical copy of dissertations so that they could be digitized using the feed scanner. Initially there were concerns about discarding the physical copies; however, these concerns were alleviated upon discovering that additional copies existed on microfilm and in print at the Collections Management Facility.

We wondered if libraries are obligated to archive dissertations. We did not come across any mention of this issue during our fact-finding phase, leading us to believe that policies regarding dissertation archiving likely vary institution by institution. For instance, Binghamton University's Graduate School once required students to provide a print copy of their dissertation to the libraries. However, this policy changed in the mid-2000s. Students are now only required to deposit a copy in ProQuest Dissertations and Theses. The libraries are not currently collecting print or electronic dissertations, and we have not encountered any evidence that the libraries are obligated to preserve the historical print dissertation collection. Although we did not investigate this issue further during this phase, we plan to conduct a more detailed examination of historical Graduate School handbooks to better understand the evolution of this policy.

Additional Resources and Expertise

The libraries already had a feed scanner that made digitizing a collection of this size efficient and effective, as well as a Zeutschel Overhead Book Scanner for scanning oversized or fragile documents. Additionally, our Preservation Department has a Challenge Titan 200BC guillotine, which made it possible to disbind dissertations in seconds. Given that we were fortunate to have access to technology and equipment to facilitate the workflow, our primary concern was whether we needed coding

expertise. For example, we had seen workflows from institutions that included Python scripts and batch editing.[5] Ultimately, we realized as we developed the workflow that such functions were unnecessary for our needs.

Metadata Requirements

Until the early 2000s, the Graduate School provided the libraries with copies of dissertations. The libraries bound these dissertations and created original Machine-Readable Cataloging (MARC) records. Subsequently, the dissertations were then added to the libraries' oversized collections. We met with the Cataloging Department to determine what metadata from the MARC records could be used for the Institutional Repository records. Ultimately, we deemed that the following MARC fields from the catalog records were relevant:

- Title (245 Title Statement)
- Author (100 Main Entry—Personal Name)
- Publication year (264 Production, Publication, Distribution, Manufacture, and Copyright Notice)
- Subject (650 Subject Added Entry—Topical Term)

The information we wanted to include in the repository records—but was not present in the MARC records—included degree name, department, advisors, and abstract. We determined that this information could be captured from the physical copy at the time of digitization.

Project Preparation

After decisions about policy and workflow, the project's next phase involved preparing for digitization and eventual ingestion into the repository. This phase involved creating a project charter that outlined, among other things, a project timeline, the roles and responsibilities of project members, project milestones, and deliverables.

Our ILS and discovery coordinator generated an Alma export of the MARC fields noted above. The Digital Commons repository uses the Dublin Core metadata schema, so converting the MARC metadata into Dublin Core was necessary. For author and keywords, this required a bit of editing. For example, the author's name in MARC was *Last Name, First Name*. This was changed to *First Name Last Name* for Dublin Core. The keywords came from the MARC 650 field. The cataloging librarian used Open Refine for data format and cleanup. They also reviewed the subject headings and ran them through MarcEdit. Subject headings like "academic dissertation" were deleted because they were repetitive and existed in almost every record. In addition, more descriptive subject headings were created. The metadata was then formatted into a metadata template to allow batch uploading into the repository.

Project Workflow

The workflow solidified after a month or so of initiating the project. Three student workers dedicated most of their spring semester, approximately 220 hours, on this project.

1. The digital initiatives assistant creates batches (e.g., lists of titles) of approximately forty dissertations, with each student responsible for a portion of the titles. The decision to split the process up into forty or so titles at a time was due to two reasons. The first is that it allows the stacks maintenance coordinator to gather a manageable amount of material at a time once every two weeks. Another rationale for breaking the project into smaller chunks is that it keeps the momentum going and allows each student to stay on task, which can be a struggle with large-scale projects.
2. The list of titles is sent to the stacks maintenance coordinator, who pulls materials, changes their physical location in Alma so patrons do not think they are available, and delivers the dissertations to the Digitization Lab.
3. The digital initiatives assistant and student workers remove the cover bindings using the guillotine in the Preservation workspace. At this time, Preservation Department employees review the items for mold or other physical condition concerns. Staff will also identify, as best as possible, which dissertations have photographs, tipped-in pages, or other additional materials held in a pocket in the binding.
4. Each student worker digitizes their assigned dissertations using the feed scanner. By using the scanner's InoTec Scamax's proprietary software, students are able to create PDF files that support optical character recognition (OCR). The created files are saved on a network drive using the following file naming convention: LastName_FirstInitial_Year (e.g., Doe_J_2024).
5. Works containing unique elements, like tipped-in photographs or folding maps in the folders, are imaged separately because they cannot be scanned using the feed scanner. For these pages or elements, we use the flatbed scanner or Zeutschel. If these elements are part of the main text, the digital initiatives assistant will combine the pages scanned on the Zuetschel with the PDF or PDFs created by the student from the feed scanner. Having the digital initiatives assistant create this PDF is partially based on access to Adobe suite products and partially to have higher quality control on the final PDF created from multiple sources.
6. Each student worker updates the metadata for their assigned dissertations to include the abstract, department, degree name, and advisors. Original metadata is also reviewed and corrected.
7. The digital initiatives assistant reviews each batch and then uploads the files and metadata to the repository. Additional files, such as folding maps, are added separately.
8. The digital initiatives assistant will perform a Google search for the title and author to determine whether a dissertation has been published elsewhere, either as a book or article. If this is the case, the digital initiatives assistant will add that information to an internal note and embargo the dissertation. The digital initiatives assistant will also embargo creative works (e.g., plays, poetry, fiction). These dissertations will be assessed at a later date to determine whether the embargo is warranted.
9. The discarded print copies are sent to recycling at the off-site storage.
10. The digital initiatives assistant sends the cataloging librarian a list of repository URLs. The cataloging librarian updates the Alma records (physical holdings records) to remove the print holdings and includes a link to the digitized copy.

Project Challenges

The project was proposed in September 2023, and digitization began in November of that year. Initially, we expected the project to last well into the summer of 2024. However, all 298 dissertations were digitized and available in the repository by June 2024.

The biggest challenge, and one that we should have anticipated, was the metadata. We aimed to include abstracts, the department's name, and the advisor's name; we believed we could simply copy this information from the OCRed PDF. As we discovered, however, very few dissertations included an abstract. Advisors' signatures were often illegible, and not all dissertations included the department name. Fortunately, one of the student workers helping with the project was a graduate student who was both creative and resourceful. He created abstracts and used his research skills to track down the names of former advisors. With the help of the university archivist, we used course catalogs from that era to track down the illegible names and verify which department advisors worked in.

We also discovered that departments were often not referred to in the dissertation, while other departments no longer existed. Our graduate student could usually deduce the name of the department based on the subject matter or the committee members (that is, once we figured out who the committee members were, which was often an ordeal on its own).

We also did not initially consider that some of the dissertations would be creative works or that some works had been published elsewhere, in part or whole, since the dissertation had been completed. Creative works are more highly protected by copyright, so if there were a misinterpretation of copyright or any other concerns, these works would be the most likely to have an issue. Some creative works still have high publication value even if they have not been published because the content may not age at the same rate as scientific research. These works were given an initial embargo period of one year for the team to research the copyright.

Initially, we planned to add all the metadata to the repository and then add the PDFs later. Once we began, it became clear that uploading the records first would be more work in the long run because of updates to the metadata. The batch updating process would have been just as time-consuming as the original metadata upload. As such, uploading a complete record with reviewed metadata in a batch made more sense.

Conclusion

We are pleased with the progress and outcomes of the project. From conception to completion, the project took approximately nine months. We attribute the project's success to the following:

- Proactive project planning
- A well-defined project charter
- Using existing resources

- Support from colleagues and administration
- A simple yet effective workflow

Key to the project's success was also the collaboration with other library departments. The Preservation Department trained the digital initiatives assistant in equipment use of the guillotine, provided access to the facilities, quarantined the few titles that had been affected by mold, and assisted with disbinding some of the oversized dissertations that were too thick for the guillotine. The Cataloging Department provided metadata cleanup as well as expertise on metadata best practices. Reader Services pulled the dissertations from the stacks and updated the holdings in Alma.

After all project materials had been digitized, the director of libraries' constituent development and the library dean sent a letter to authors whose works had been digitized to highlight the project's goals and accomplishments. The letter also emphasized the libraries' commitment to supporting scholarship, research, and open access initiatives. We have already received several enthusiastic responses from alumni authors. An article about the project was also published in the campus newsletter, *BingUNews*.

Because the project was only recently completed, we have not been able to conduct an assessment on the number of downloads. Moving forward, we would like to meet with the Graduate School to discuss the possibility of digitizing more recent dissertations. As part of this conversation, we would like to discuss how we can collectively encourage current students to deposit their dissertations in the repository in addition to ProQuest.

Acknowledgments

We would like to acknowledge the following people who were instrumental in the project's completion: Sasha Frizzell, Heather Parks, John Lee, Tina Burrows, Isiah Coombes, Lea Son, and Carrie Blabac-Myers.

References

1. Gail Clement and Melissa Levine, "Copyright and Publication Status of Pre-1978 Dissertations: A Content Analysis Approach," *portal: Libraries and the Academy*, 11, no. 3 (2011): 813–29, https://doi.org/10.1353/pla.2011.0031.
2. Daniel S. Dotson, "Analysis of Usage of The Ohio State University's Electronic Theses and Dissertations," *Library Philosophy and Practice* (2019): 1–13, https://digitalcommons.unl.edu/libphilprac/2486/.
3. Meghan Banach Bergin, "Providing Online Access to Over a Century of Theses and Dissertations at UMass Amherst," presentation given at the 2019 Northeast Institutional Repository Day, June 18, 2019, https://repository.escholarship.umassmed.edu/handle/20.500.14038/37434?show=full; Christy L. M. Shorey, "Engaging Alumni: The How and Why of Author Outreach for Dissertation Scanning Projects," presentation given at the Charleston Library Conference, Charleston, SC, November 2018, https://doi.org/10.5703/1288284317071.
4. Clement and Levine, "Copyright and Publication Status of Pre-1978 Dissertations."

5. Jeremy Moore and Jason Long, "A Novel Workflow for Large Scale Thesis Digitization," presentation at the Texas Conference on Digital Libraries, Austin, TX, May 2016, https://hdl.handle.net/10877/16516; Kathryn Loafman, Daniel Gelaw Alemneh, and Jeremy Berg, "From Theses and Dissertations to ETD: Retrospective Digitization and New Forms of Scholarship," presentation at the 2014 CTLC Conference, Dallas, TX, July 25, 2014, https://digital.library.unt.edu/ark:/67531/metadc333035/.