

From Paper to Pixels: Rough Spots and Roadblocks on the Way to ETDs

Max Read

Communications and Thesis Coordinator, Faculty of Graduate Studies
The University of British Columbia, Canada
max.read@ubc.ca

Bronwen Sprout

Digital Initiatives Librarian, UBC Library
The University of British Columbia, Canada
bronwen.sprout@ubc.ca

This paper describes the process of developing an ETD program from pilot project to mandatory ETD submission, and will discuss issues and solutions developed by the UBC Library and the Faculty of Graduate Studies.

In December 2005 the UBC Library and Faculty of Graduate Studies initiated a pilot project for ETD submission to the Library's new electronic repository. We spent several months researching and preparing before proceeding.

When ETD submission opened in November 2007 we immediately had 50% of theses submitted electronically, and that percentage has increased. Theses submitted on paper are scanned to PDF and handled the same way as ETDs.

The most challenging issue was the question of archiving ETDs. We decided to discontinue the older technology of microfilm/microfiche/paper backups, and to put resources into developing a reliable system for electronic archiving. The process leading to this decision involved both the Library and the Faculty of Graduate Studies, though each had slightly different reasons for the choice.

Some unexpected issues arose: creative arts faculty and students did not want their theses online, and we had an increase in requests to withhold theses on the grounds that journals would not

publish previously-online material. These issues have been resolved, albeit temporarily, and steps are being taken to find permanent solutions.

Technical support is now available to students through a Library computer lab, and we will be proposing that electronic submission become mandatory. This will require additional streamlining of the submission processes wherever possible, and there are various reasons for the delays in implementing them.

Thesis Submission and Archiving: The Dark Ages

Before the UBC ETD project, the Faculty of Graduate Studies reviewed and accepted paper theses in person and shipped them for archiving on behalf of the Library. Doctoral dissertations were sent to ProQuest Dissertations and Theses to be filmed, fished, and scanned to PDF. It took a minimum of one year from submission date for full text dissertations to be available in the ProQuest database. Further, ProQuest's arrangement with Library and Archives Canada (LAC) meant that ProQuest distributed the dissertations for three years before making them available in LAC's Theses Canada Portal. Master's theses were microfished locally, and were only available through the UBC Library. The sole reason for this was the difference in cost: \$39 per dissertation at ProQuest as opposed to \$9 per thesis for local fishing. (Although ProQuest does more than just microfiche the thesis, the budget wasn't available to cover the additional cost.)

The Pilot Project

In November 2005 the Library approached Graduate Studies with the suggestion of carrying out a pilot project for ETDs. The purpose of the project was to develop a system that would provide immediate online access to accepted UBC theses and dissertations through UBC and through the LAC Theses Canada Portal. At the time the Library was in the early stages of developing its Information Repository and already had DSpace software installed.

Planning and preparation

The ETD team investigated procedures at other universities to help make decisions about policies, workflow, and DSpace customizations. We spoke to library colleagues at the University of Manitoba and obtained temporary access to their submission process to see what DSpace customizations they had made, and corresponded with the University of Victoria and studied their project proposal and results report.

As soon as the pilot project was approved, we developed an informational web site about ETDs, and used this to help promote the project.

Graduate Studies consulted UBC's legal counsel and obtained wording from them for the UBC Thesis Licence so that students were clearly agreeing to have their theses available online.

The Library identified the metadata needed for the ETDs (based on ETD-MS) and customized Dspace accordingly. To reduce error in the student-entered metadata the Faculty of Graduate Studies recommended setting drop-down menus in DSpace for degree name, program name, and graduation date. When students are asked for the name of their degree program they're sometimes very creative!

Execution

The scope of the project was to accept 30 theses online from a variety of programs across campus. The system was ready to accept theses at the end of August, and we had met our target of 30 ETDs by the end of October.

Students submitted their ETDs to DSpace, and added metadata as part of the submission process. The thesis staff at the Faculty of Graduate Studies were automatically notified by e-mail when a thesis was waiting to be reviewed, and a staff member approved or rejected the thesis. If rejected, the student was informed by e-mail that he/she must make corrections and re-submit the thesis; this e-mail included detailed instructions about the required corrections. This process continued until the thesis was correct and accepted, and is almost the same as the process used today. During the pilot, the Library authorized log-ins for students based on requests by the Faculty of Graduate Studies; today, this part of the process has been streamlined and the Faculty of Graduate Studies authorizes students. Additionally, post-pilot project the Library edits student-submitted metadata. Essentially this submission and approval process mirrored the existing process in an online environment, and was remarkably trouble-free.

Concluding the pilot project

We did not develop a separate system for archiving the pilot ETDs, so they were printed out at the Faculty of Graduate Studies and sent through the usual channels to ProQuest and Micro Com.

Feedback from students was overwhelmingly positive (with one notable and highly eccentric exception); the majority agreed that the instructions provided were good, the in-person/phone help was useful, and the system for submission was straightforward.

Post-Pilot Developments

Given the success of the pilot project, the ETD team received approval from the UBC Library and the Faculty of Graduate Studies to move ahead to make electronic thesis submission a permanent option for students, with the goal of making it mandatory in the future. This required further development and decisions in order to streamline the process as much as possible.

DSpace

The Library created communities within DSpace (now called cIRcle, UBC's Information Repository) for different types of submissions, and established a collection policy for cIRcle. The Library's cataloguing staff agreed to check the metadata for each ETD prior to accepting it into the collection.

Library and Archives Canada can harvest ETDs automatically, using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). LAC worked with the UBC Library to set up and test the harvesting procedure.

Archiving ETDs

Before accepting any more ETDs we needed to review our archiving practices and develop a system of purely electronic archiving.

The development of an ETD program meant that the advantage of making theses available in the ProQuest database was reduced. ETDs are quickly available and searchable by other means, and Library and Archives Canada's Theses Canada Portal is a well-known resource for Canadian

theses. (However, it should be noted that ProQuest does provide a copyediting service and checks for copyright violations, in addition to creating an archival microfilm backup.)

We considered three options for archiving ETDs:

1) Local electronic archiving: ensure that the PDF files are stored in two locations, on separate servers. This means that a backup copy would be quickly available locally. Another copy will be kept by Library and Archives Canada, and in the event of a disaster, the multiple copies of electronic documents will ensure safe backup.

2) Microfilm: send PDF files electronically to be converted to microfilm backup. Not many companies can make microfilm directly from a PDF due to the high cost of the machine, and the obvious choice was to send both doctoral and master's theses to ProQuest. Unfortunately the cost of this far exceeds our budget.

3) Status quo: accept and distribute theses electronically and continue to send doctoral theses to ProQuest for archiving and master's thesis to the local company to be microfiched. This means that if the electronic copy of a master's thesis were destroyed, the backup would only be on microfiche. This option was not recommended by the ETD team.

The UBC Graduate Policy Committee was in favour of the first method—all-electronic archiving of electronic theses—and this was the method adopted. In addition, the UBC Library is participating in a LOCKSS (Lots of Copies Keep Stuff Safe) initiative (the COPPUL Private LOCKSS Network) and the ETD collection is under consideration for archiving by the network. Once adopted, UBC's ETDs will be stored on many servers across Canada.

Because electronic submission is not mandatory, a number of theses are still submitted in paper format. The Library advised that archiving electronic theses and paper theses by different methods (electronic archive and microfiche with paper backup) would cause confusion when searching for theses at UBC. Consequently, theses submitted on paper are now scanned, an OCR'd PDF file is created by the Library, and the PDFs are submitted to cIRcle by Faculty of Graduate Studies staff. Thus all UBC theses and dissertations are now available online shortly after submission, whether submitted as ETDs or on paper.

Ongoing Developments and Issues Arising

In November 2007 electronic submission of theses was opened to students. Immediately about 50% of theses were submitted electronically, which is far more than we'd anticipated. This percentage has increased, and so far in 2009 65% have been electronic submissions. Interestingly, just before each thesis submission deadline, more theses are submitted on paper than electronically. This appears to be because of the slight delay between students submitting their approval and licence forms and setting up an account for submission to cIRcle; when time is tight, students print the thesis and bring it to the office in order to complete the process more quickly.

We are continuing to streamline and improve the process of submission and handling of ETDs at UBC. Some of the situations and solutions we have encountered are described below.

Withholding ETDs from publication

We needed to develop a method of withholding ETDs from publication to give students and the UBC Industry Liaison Office time to file for patents, sort out intellectual property issues, and so

on. The number of requests for theses to be withheld increased once all theses become quickly available online, as some faculty and students assumed that journals would not publish previously-online material. Some of these requests were withdrawn when a little research demonstrated that this was not the case.

Ideally, we'd like to be able to direct restricted theses into a separate collection in cIRcle that is accessible only to cIRcle administrators or to people specifically granted access, but at the moment the Library's limited programming resources do not allow for this development. Initially we required paper submissions from students who requested their theses be withheld, but recently we developed a workaround that enables us to withhold electronic theses. Students whose theses have been approved for withholding add "WITHHOLD" in front of their title. When we see this in DSpace, we log out and log in again with a separate e-person identity. The thesis is reviewed and approved by this identity, and is kept in this separate "Owned Task" section during the withhold period. This is far from a permanent solution, but until the Library develops a system so that theses can be held there, it's at least allowing us to withhold ETDs.

Multimedia

Previously, students submitting paper theses could also submit accompanying materials—for example, additional data on a CD-ROM, music or video files on a CD or DVD. These were catalogued, and would be signed out with the microfiche copy. At the moment the Library doesn't have a system for connecting these accompanying materials to an electronic thesis, although they plan to develop one. In the meantime, students are submitting their accompanying materials to the Faculty of Graduate Studies, so they can be passed on to the Library when the system is able to handle them. Eventually, the metadata record will likely be edited to note that there is accompanying material and students will be directed to the Library to listen to/view the material.

In theory, these multimedia files could be embedded in the PDF. However, at present Library and Archives Canada is only able to harvest electronic theses as single file PDFs with no multimedia components. They are working to remove this limitation, but in the meantime, as it is very important to UBC to have its theses in LAC's Theses Canada Portal, their limitations are also ours.

However, we have experimented with workarounds to these limitations, including the recent thesis submission in which a student reported his examination of ways to assess the stability of water flow channels in steep streams. He included video clips of time-lapse photography that showed the patterns of water flow over various types of stone streambeds constructed in the lab. It would require a lengthy description to communicate what is shown in a few seconds in the video clips. We asked the student to submit a copy to cIRcle without the video clips, and to include a link to a web site where the complete thesis can be downloaded. We have kept a copy of the complete multimedia thesis so that we can substitute it for the other once the LAC limitation is removed.

Creative Arts theses

In March 2008 we were approached by a delegation of faculty and students from the Creative Arts disciplines. They were concerned that ETD online publication would create obstacles to later press publication / distribution for creative writing, film, and music (both scores and recordings of performances). They wanted creative arts students to be able to have their theses permanently

withheld from online publication, and instead have only one copy available through the UBC Library.

The Faculty of Graduate Studies proposed an interim solution for those students who came into the program before theses were made available online. Students would submit a paper copy of the title page, abstract, and table of contents of their theses, and would also submit a PDF of the entire thesis on a CD-ROM, CD, or DVD. The introductory material would be digitized and submitted to cIRcle, but the digital media would be available only on loan from the UBC Library.

Creative Arts agreed to this, and a special Thesis Licence has been drawn up which enables students to agree to having their preliminary pages online, but which also affirms that the accompanying digital media will only be available at the UBC Library.

However, UBC's institutional repository is meant to include actual content (rather than to be used as a catalogue for materials that actually exist elsewhere), so theses submitted under the Creative Arts Thesis Licence are sitting in a box at the Faculty of Graduate Studies until we decide what, if anything, can be done with them.

Technical assistance for students

Technical help with PDF creation recently became available to students through the Chapman Learning Commons at the UBC Library. Now that students have access to both the software and the assistance they need, we are developing a proposal to make electronic thesis submission mandatory for all graduate students.

Paperwork management

We need to streamline the processes of collecting committee approvals of a thesis and of distributing receipts once the thesis has been accepted. At the moment UBC Legal Counsel is not in favour of anything other than original signatures, but at some point we hope to set up a web-based system for committees and examiners to give final approval on theses, and to automate a system for providing a thesis receipt.

Retroactive digitization of theses and dissertations

The UBC Library is preparing to digitize all paper and microfiche theses in-house. Since the 1920s graduate students at UBC have generated approximately 33,500 theses totalling just under 5 million pages. The Library hopes to be able to provide full-text access to as many older theses as possible. A pilot project undertaken in the summer of 2008 resulted in the digitization of approximately 1,000 theses from 1993. Since that time the UBC Library has begun a project to include theses produced between 1992 and 2007. Currently there are approximately 5,300 retrospective titles available in cIRcle with more being added every day. PDF copies of the theses have been secured such that text and images may not be cut and pasted into new electronic documents. If authors have any concerns about having their theses digitized and included as part of this collection they can notify the Library and the thesis will be immediately removed.

Thesis Submission and Archiving: The Age of Light

Together, the ETD program and the retrospective thesis project promote open and comprehensive access to a significant body of unique information created by UBC graduate students. From the UBC Library's point of view, these projects hugely improve access to UBC theses for Library users: when the retrospective project is complete, not only will all UBC theses be searchable in

one database, but all theses will be available in full text. From the point of view of the Faculty of Graduate Studies, our students' research is now quickly and easily accessible, promoting their work and easing collaboration between researchers. Students who submit electronically save paper, money, and time, and can submit from off-campus. At the moment there are costs associated with scanning the paper submissions, but once electronic submission is mandatory we will realize considerable cost savings.

The success of the ETD program can be attributed in large part to the strong partnership between the Library and the Faculty of Graduate Studies. In future, as we move towards mandatory online thesis submission, we anticipate that the strength of this partnership will continue to ensure project success and meet the goals of both partners by simplifying the thesis submission and approval process and improving access to UBC theses.