Study Leave Report
April – September 2009

Submitted to Marnie Swanson, University Librarian
by Nancy Stuart
Oct. 26, 2009

The goal of the study leave was to increase and share my knowledge and expertise in the area of electronic theses and dissertations. I am also interested in the goal of the Theses Canada program to provide leadership and support to Canadian universities with respect to their theses programs. I would like to take a leading role in moving forward ETD programs in Canadian universities and assisting in the development of the Theses Canada Portal and its services.

There were four areas of interest I wished to explore in more detail. The first was an outreach program to various Canadian institutions in order to facilitate their transition from print to electronic theses. The second was to forge a stronger link between CARL Institutional Repositories program and the Theses Canada program. The third was to study an emerging standard for the harvesting of multimedia/multiple file ETDs. The final area was to create and disseminate "Best practices" in the area of ETDs.

To accomplish the goals of my study leave, I secured a placement at Library and Archives Canada, Ottawa to collaborate with Ms. Sharon Reeves, Manager of the Theses Canada program. A study space was provided by Theses Canada and I was given a security pass for the LAC Building at Place de la Cité, Gatineau. I was also provided with a workstation, access to the LAC intranet, an email account, and a telephone. I would like to acknowledge the support I was given by Theses Canada and LAC for the generous use of their facilities.

OVERVIEW

The study leave goals formed a structure for the tasks to be accomplished within the 6 month time frame. The goal of outreach to the Canadian institutions was accomplished in the following ways. A survey was conducted among 50 Canadian institutions. Also a Canadian ETD and Open Repositories Workshop is in the planning process for May 2010, sponsored jointly by LAC, Carleton University and University of Ottawa. A meeting with Kathleen Schearer, Manager of the CARL Institutional Repositories program was held to forge closer links with the CARL IR program and the Theses Canada program. Upon further investigation into ORE standards, it was determined that it is too soon for the Theses Canada program to move forward on the harvesting of multimedia/multiple files. The final goal of creating ETD workflow and

metadata best practices is on-going, but a number of documents were produced during my study leave to support best practices in the area of ETDs.

OUTREACH

Theses Canada had conducted a survey of the Canadian institutions that belonged to the Theses Canada Program in 2005 with an update in 2007 to determine the status of their ETD submission programs. At that time only 10 universities had ETD submission programs and only 5 universities were being harvested by LAC: Waterloo, Laval, Saskatchewan, Manitoba and Queen's. In 2008 another 3 libraries had joined the harvesting program: McGill, Victoria, and UBC. Another survey on ETD initiatives at Canadian institutions was due and I undertook it as my first task of my study leave.

This current survey for Theses Canada was conducted in April/May 2009 by telephone or email to 42 institutions across Canada. Information was also gathered from previous surveys on 8 institutions. Most of the information was gathered through telephone conversation and supplemented by information gathered via email. This survey gives a flavour of the state of IRs and ETD submission programs in 50 Canadian institutions as of May 2009. The survey document is included in Appendix A.

This survey complements two other surveys that were conducted by the Ontario Council of University Libraries (OCUL) E-Theses Task Force in the Fall of 2008 and by the Canadian Association of Research Libraries (CARL) IR Program in the Spring 2009. The IR and ETD environment is rapidly changing as more institutions implement Institutional Repositories and begin ETD submission programs. The goal of the Theses Canada Program of Library and Archives Canada is to build a digital collection of Canadian theses and dissertations. Currently LAC's digital theses and dissertation collection stands at 83,000 and is expected to grow exponentially over the coming years.

The results report of the survey was produced including the raw data and compiled statistics. It was distributed to all participating institutions in August 2009. Theses Canada plans to put the results report on the Theses Canada website. The results of this survey have proven to be an impetus for institutions to move forward on their ETD programs. 2010 should prove to be a busy year for ETDs in Canada.

Another aspect of outreach to the Canadian universities was to organize an ETD workshop. With my help, Sharon Reeves, Manager of the Theses Canada Program, was able to form an organizing committee to begin planning the Canadian ETD and Open Repositories Workshop. Carleton University, University of Ottawa, Theses Canada and I formed the organizing committee. We are in the planning stages of the workshop to be held at Carlton University, May 10-11, 2010. A public announcement about the workshop was distributed Sept. 30, 2009 and a workshop website has been set up. I am the website manager, using Open Conference Software hosted at University of Victoria.

I will act as the Registrar for the two day workshop and help with other tasks associated with the workshop over the next 6 months.

COLLABORATION

The goal to network more with the CARL Institutional Repositories program was realized. Sharon Reeves and I met with Kathleen Schearer, Research Associate at CARL, in Ottawa to discuss ways of collaborating and creating a stronger bond between the CARL IR program and Theses Canada program. Kathleen agreed to do a presentation about the status of IRs in Canada at our ETD and Open Repositories Workshop next spring. The CARL IR Program was added as an organizational member of the Theses Canada Advisory Committee. Each program agreed to link to the other on their own websites as a way of promoting the other's program.

Both programs conduct a survey of Canadian institutions every other year. There was agreement to collaborate in future surveys as many of the questions overlap and each program is interested in seeing the results of the other's survey.

BEST PRACTICES

The first formal ETD program at a Canadian university began in 1999. Over the past ten years ETD programs have expanded and grown. Often there was little institutional support for an IR or an ETD program and decisions were based on the available technology at the time. Institutional Repositories and the digital content they archive are one of the fastest changing areas in the digital landscape. This applies to both technology and standards and has implications for both IR and ETD policies and practices. There is definitely a need for some "best practices" documents in these areas.

The Theses Canada website has documentation on "How to set up an ETD program". This document was in need of updating. One of my first tasks was to rewrite the document with careful attention to include Canadian examples and current practice. The updated document is included in Appendix B. I drew heavily on my own experience of setting up an ETD program at UVic. Over the years I have attended 4 ETD conferences. I attended ETD 2004 in Louisville, KY, ETD 2006 in Québec City, ETD 2008 in Aberdeen, Scotland and ETD 2009 in Pittsburgh, PA. At these conferences I have learned many best practices and networked with people involved in ETDs all over the world. It was at ETD 2004 that I met Sharon Reeves as we were the only two Canadians in attendance at that conference. With the knowledge gained at the ETD conferences and the support of the Theses Canada program, UVic was able to become the 10th Canadian university to set up an ETD program and the 7th to have our ETDs harvested by Library and Archives Canada.

It has been a wonderful opportunity to work and collaborate directly with the Theses Canada program and the IT staff at Library and Archives Canada. This one on one interaction over a period of six months has been most fruitful. I have gained a better understanding of the Theses Canada program and the operational aspects of the LAC harvester. Both Sharon and the IT staff have been very generous with their time. The knowledge gained has helped me to create metadata and workflow best practices.

In the process of creating a Best Practices document for ETD metadata and workflow, I consulted with professionals in institutions across Canada with established ETD programs. The final ETD workflow and metadata best practices document will be presented as a 1 ½ hour session at the Canadian ETD and Open Repositories Workshop May 10, 2010. Over the next six months I will consult with various universities to compile a set of best practices that cover such topics as: IR selection, file formats and naming conventions, mandatory submission policies, metadata fields and schemas, ETD websites, ETD guidelines, harvesting standards, restricting theses, copyright and preservation issues. The finished document will be published in UVicDSpace after May 11, 2010.

ORE

During my study leave I read a file on ORE (OAI Object Reuse and Exchange) – a standard to represent and exchange complex digital objects in a web-based environment. ETDs have opened the door for complex digital objects. There definitely is a need to clearly describe how the parts of a digital object relate to each other. ORE is an emerging standard that addresses this issue. The conclusion I drew is that this standard will be useful in resource discovery of aggregations of web resources. But, currently the ETD-MS metadata that is harvested by the LAC harvester lacks any structural metadata that would describe an object with multiple files. Therefore LAC only harvests a single PDF file. ORE could be a long term solution for multiple file harvesting or ingesting, but as yet it is a very new standard with the first version being released in Oct. 2008. In the interim, Adobe may have a solution for wrapping up multiple files into a single PDF file.

My attendance at the ETD 2009 conference in Pittsburgh augmented my knowledge and expertise, especially in the area of dealing with multiple files. I attended a 3 hour pre-conference workshop on Adobe Acrobat 9. The new feature in Adobe Acrobat 9 has the ability to bring multiple files together in a single organized and interactive PDF Portfolio. This ability will be useful for multimedia and multiple file ETDs. While on study leave I was not able to trial this product, but on my return to UVic I will experiment with it using some of our multiple file ETDs. If the experiments are successful, I plan to upgrade our UVic multimedia and multiple file ETDs using this product. The LAC harvester will then be able to harvest all ETD content from UVic.

Adobe 9 has the capability of bringing together PDFs, excel files, datasets, audio files, and video files together into one neat PDF file. Currently these supplementary files are not harvested by LAC, only the main PDF file is harvested. This has huge implications for preservation as well.

CONCLUSION

My 6 months at Library and Archives Canada also presented some unique opportunities. Sharon Reeves and I were able to meet with the metadata standards librarian, who sits on the MODS (Metadata Object Description Schema) Editorial Committee, and discuss some of the challenging issues around metadata schemas and ETDs. These discussions were very helpful in clarifying issues we deal with daily in managing an ETD collection, whether at an individual institution or LAC's digital theses collection. The information was also useful for incorporation into the "best practices" presentation, which I am currently developing.

My learning from this study leave will definitely contribute to my ability to meet the changing needs of ETDs at the University of Victoria and across Canada. I will be able to contribute "best practices" both in the workflow in the Faculties of Graduate Studies and the workflow of ETDs in Libraries and Institutional Repositories. I have gained a much greater understanding of the ETD workflow from start to finish; in particular the technical requirements for ETDs to be harvested by the Library and Archives Canada harvester. I have a much better appreciation of the complexities of metadata schemas and metadata crosswalks, as I spent a great deal of my time studying them. I look forward to the challenges ahead as ETD collections across Canada expand and grow over the next decade.

APPENDIX A

ETD Initiatives at Canadian Universities
Summary of Survey Results
July 2009
Updated Sept. 2009

by Nancy Stuart
for Theses Canada
Library and Archives Canada

The survey was conducted in April/May 2009 by telephone or email to 42 institutions across Canada. Information was also gathered from previous surveys on 8 institutions. The responses from the 50 institutions are listed at the end of the survey along with the survey questions. Most of the information was gathered through conversation and at the time of reading the information may have changed. This survey gives a flavour of the state of IRs and ETD submission programs in Canada as of May 2009.

In Canada, 16 institutions (32%) have implemented a student submitted ETD program. The first ETD program began at the University of Waterloo in 1999, and over the past 10 years another 15 universities have implemented an ETD program, with 5 implementing in 2008/09. All 16 ETD submission programs (100%) were requiring the ETD to be in PDF format. Some institutions also accept other file formats, such as html and xml, and for supplementary material formats such as Excel, Jpeg, Mpeg, Gif, etc.

The most popular software used was DSpace. Others used were Archimède, DigiTool, ContentDM and ETD-db and EPrints. Some institutions in the early implementation stages of their IR are using Fedora and Digital Commons.

Fourteen institutions (88%) send their ETDs to ProQuest, while 2 have gone paperless and have only ETDs and have stopped sending their ETDs to ProQuest. Of the 50 institutions that belong to the Theses Canada program 48 continue to send their theses to ProQuest under the current contract.

ETD submission is mandated at 7 institutions (44%) while the rest have a voluntary program. Some institutions are hoping to begin their ETD program as mandatory from the beginning. It is the experience of some that it is easier to have one program, electronic, than to run two programs, one for print and one for electronic.

Currently 11 institutions (69%) are being harvested by LAC, 2 will be harvested in the very near future and 3 are not yet ready to be harvested. Of the

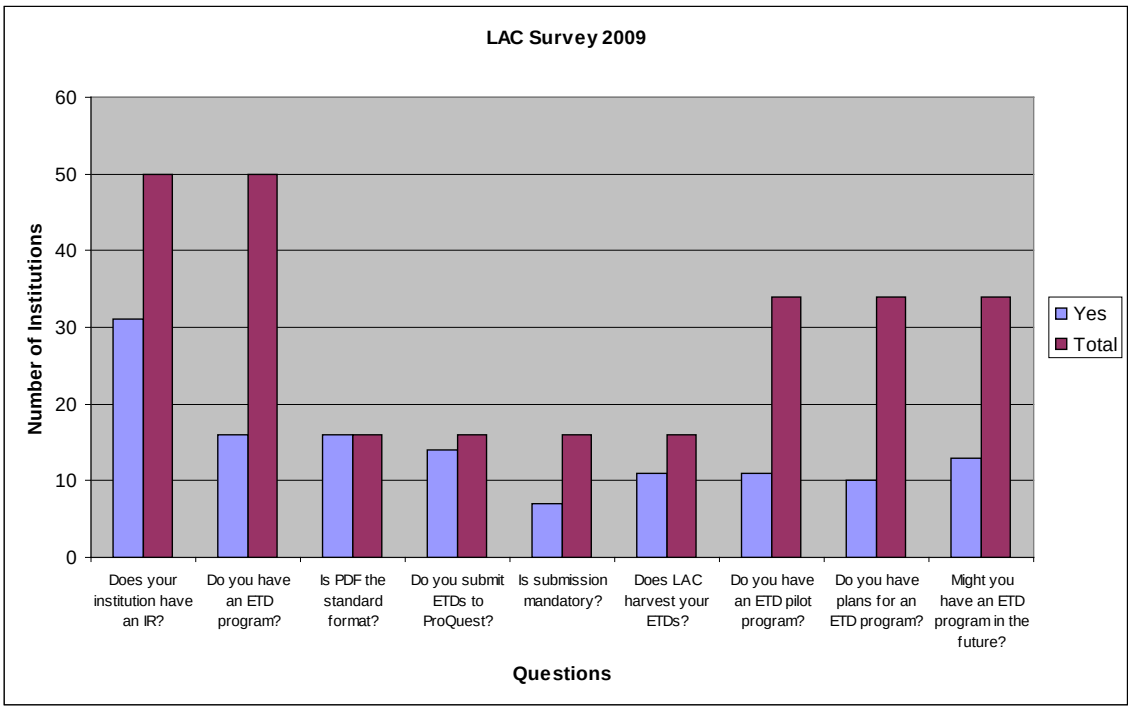
50 institutions, 31 (62%) have already implemented an Institutional Repository, 19 do not have an IR.

Of the 34 institutions that do not have an ETD submission program, 11 (32%) have an ETD pilot program in progress. Ten institutions (29%) are at the discussion stage of beginning an ETD program and 13 institutions (38%) would like to have an ETD program, but there are no formal plans in place.

During the telephone survey, many institutions had questions about sending their ETDs to ProQuest. Many are overwhelmed with the handling of print theses, so want an easy and efficient way to send their ETDs to ProQuest. The suggestion was to have the students submit a PDF copy on a CD-ROM. The institution can then send the CD-ROM along with the paper work to ProQuest. Another option is to FTP the electronic copy. This takes more IT staff support and a special arrangement with ProQuest.

Other questions related to the setting up of the program. Library and Archive Canada is in the process of updating the content on the Theses Canada Portal and the first document to be updated is the How to Set up an ETD Submission Program. Other questions were in the area of requirements for harvesting. The General and Technical Requirements document will also be updated in the near future.

Library and Archives Canada, along with Carleton University and University of Ottawa are planning to hold the Canadian Open Repository and ETD Workshop to be held May 10-11, 2010 at Carleton University.



Survey on ETD Initiatives at Canadian Universities
2009
conducted by telephone

Does your university have an electronic theses submission program in place?

If yes it has an ETD program:

How long has it been in place?

What formats are acceptable for electronic theses and accompanying material, e.g. PDF, XML, Excel, Media formats, etc.?

What software did you use for the ETD submission program, e.g. ETD_db software from Virginia Tech, Dspace, Fedora?

Are you submitting ETDs to Theses Canada via ProQuest Information and Learning as well?

If so, how? Via FTP; PDF on CD-ROM, paper

Is student participation voluntary or mandatory?
Do you anticipate implementing a mandatory submission program?
When?

Are you ready to be harvested by LAC?

Do you create MARC records for your catalogue from the DC metadata?
Would you like to?

If no it doesn't have an ETD program:

Does your university plan to set up a program to accept electronic theses and dissertations from students?

If no, say thanks and ask them to contact Theses Canada in the future if the university changes its plans.

If yes, please identify what stage you are at:

a) informal discussions have taken place but an ETD program has not been approved

b) formal committee is in place and is developing a proposal for approval

c) approval has been given to conduct a pilot project and planning for the pilot is underway

d) a pilot project is in progress

If you are planning/conducting a pilot project:

a) In what format will/do you require students to submit ETDs and supplementary material, e.g. PDF, XML, Excel, Media formats?

b) What software are you considering/did you select for the ETD submission program, e.g. ETD_db software from Virginia Tech, DSpace, Fedora?

c) When do you expect to finish your pilot project and implement a formal ETD submission program?

Are you aware of the Information for Universities on LACs Theses Canada Portal?

How could we help you in your transition to ETDs? Info, Personal visit

Would you be interested in attending a Canadian ETD symposium in May 2010?

APPENDIX B

How to Set Up an ETD Submission Program

The development of an ETD submission program is not the responsibility of a single group in the university. It requires the joint and cooperative work of the library staff, the IT team, the Faculty of Graduate Studies and participation of the university's graduate students.

To successfully implement an ETD program it is important to get high-level support from your university administration. A good way to do this is by writing a proposal to do a pilot project and submitting it to the appropriate senior administrative officers. To get an idea as to what the proposal should include consult the following resources to get you started.

Virginia Tech ETD Guide
University of Waterloo Business Plan, 1999
University of Victoria Project Charter, 2005

Pilot projects allow universities to put in place the proper infrastructure and procedures and to communicate the changes to the university community. Communicating the benefits clearly and frequently is an important step. Many universities have identified changing the culture at their institutions as the most difficult challenge in establishing an ETD program. The online resources that are now available from various Canadian universities (hyperlink to ETD websites.doc) that have already implemented successful ETD submission programs can shorter the implementation time and make for a smoother process.

Steps in Setting Up an ETD Submission Program

1. Preparation

It's a good idea to start by reviewing the information on a few websites and in various documents.

Virginia Polytechnic Institute and State University, http://etd.vt.edu
Networked Digital Library of Theses and Dissertations (NDLTD), www.ndltd.org ,
Canadian Universities (hyperlink to ETD websites.doc)

Another good source of information is the publication entitled Electronic Theses and Dissertations: a Sourcebook for Educators, Students and Librarians, edited by Edward Fox, Shahrooz Feizabadi, Joseph M. Moxley and Christian R. Weisser. ISBN 978-0824709730

You can also find information in the following:

Bailey, Charles W. Electronic Theses and Dissertations Bibliography, ver. 4, 2009.

2. Set up the Team

Set up a project team with representatives from the library, the IT department, the Faculty of Graduate Studies, senior administrators from the university, and the Graduate Students Association.

3. Prepare a proposal

Prepare a pilot project proposal and submit the proposal for approval. This is the stage to work out policies specific to your university. You may want to include information on publication potential, intellectual property and rights management, plagiarism, orientation and training, standards, costs, archiving and preservation and restrictions on access.

Preservation is a very important issue for ETDs. This issue should not be overlooked. One way to ensure preservation is to use the LOCKSS system. Another way is for your institution to join the MetaArchive Cooperative.

4. Software

This is also the time to decide on the technical infrastructure you plan to use. Please make sure that you consult Theses Canada requirements for harvesting your metadata and theses prior to installing your chosen software.

These days most universities are implementing institutional repositories (IRs) and including ETDs in them. The Canadian Association of Research Libraries (CARL) sponsors an institutional repositories program for its members. Information on the CARL program can be located at http://www.carl-abrc.ca/projects/institutional_repositories/institutional_repositories-e.html

There are any number of choices of institutional repository software available, both open source and proprietary. By far the most frequently implemented IR software in Canada is DSpace, which is an open source product.

Since DSpace does not support ETD-MS, the metadata standard for electronic theses and dissertations, Jonathan Roby, a programmer at the University of Manitoba, developed a Dublin Core to ETD-MS crosswalk for DSpace, which can be accessed on the University of Manitoba website at http://mspace.lib.umanitoba.ca/ETDMS.zip. This crosswalk will need to be customized for your own Institutional Repository.

Systems staff at the Edinburgh University Library have developed an add-on pack that provides the ability within DSpace to operate a supervised authoring

facility, an addition to the DSpace administration area to manage the supervising groups and their access policies to the student's work and two submission interfaces, one for ETDs and one for e-prints. Information on the add-on pack can be located at www.thesesalive.ac.uk/dsp_home.shtml.

A number of other IR software applications are available. Open source IRs include, among others, Archimède (Laval University), Eprints and Fedora. Proprietary IR software applications include Ex Libris' DigiTool and Digital Commons, both of which have been implemented by individual universities in Canada. Digital Commons provides a total end to end submission package similar to the add-on pack for DSpace. A few universities are using CONTENTdm. Smaller institutions where there is no IR may choose a hosted option, Open Repository through BioMed Central.

Non IR options are to install the ETD_db freeware from Virginia Tech or VALET for ETDs from VTLS, which is open source and designed to be used with Fedora. Both also offer a total end to end ETD submission package.

5. Create an ETD website

Once the project is approved, set up a website for ETDs. This can be done by the IT staff at your University or by the University Library or Faculty of Graduate Studies. Information on the website should include an overview of your ETD program, submission guidelines, ETD procedures, policies, information on tutorials, etc.

Check out some universities with ETD websites Don't reinvent the wheel.
Canadian universities (hyperlink to ETD websites.doc).
Virginia Tech
West Virginia University

6. Training for grad students

Implement a training program for graduate students. They will need training on all aspects of ETDs, from creating the word document, converting it to a PDF and submitting it to the IR. Some universities, such as the University of Waterloo and Virginia Tech, offer in-person training sessions. The University of Victoria has developed an online tutorial How to submit an ETD. Ohio State University has a tutorial on creating a PDF. West Virginia also has a Convert to PDF tutorial. University of Saskatchewan has created some ETD Word Template tutorials.

7. Set up the ETD workflow

Establish the ETD workflow, i.e. who is responsible for what. In the most common scenario the students upload their theses files to the Grad Studies office where they are reviewed, approved, then released to the Library. The metadata

is then made available for harvesting by other organizations such as Library and Archives Canada and the NDLTD. LAC also harvests the theses themselves (in PDF format). Here is a typical Canadian ETD workflow (hyperlink to ETD Workflow Diagram.doc). Another example is the workflow from Virginia Tech.

8. Pilot Project

Run a pilot project either with a limited number of students or with one or two specific departments. This will allow you to fine-tune your procedures and workflow. As few as 20 or 30 ETDs is sufficient to run a pilot project.

At the end of the pilot project many universities adopt a voluntary e-theses submission model for several years before moving to mandatory submission of electronic theses and dissertations. If possible recommend mandating ETDs at your institution as soon as you begin your ETD submission program.

9. Prepare for harvesting by LAC

Once your ETD submission program is established, have your IT staff prepare your ETD collection for harvesting. The IT staff need to read the Technical Requirements in order to implement the IR as an OAI data repository. There must be only one ETD Collection set. LAC requires metadata in two formats: DC and ETD-ms. After following all the technical requirements, the IT staff will need to validate the oai_etdms records. This is important to do before testing with LAC. Once your records are validated, contact LAC that you are ready to test for harvesting. LAC has limited IT staff and it is encumbant on the institutions to follow the requirements and rigourously verify their XML and OAI compliance.

10. Evaluate and enhance

Like any program, your institution's ETD submission program should be periodically evaluated and enhanced.

- Software needs to be upgraded
- Graduate Studies staff need to continually be aware of policy changes and how copyright impacts the program
- IT staff need to monitor appropriate list-servs in order to become aware of upgrades and enhancements
- Library staff can implement enhancements to the program, such as creating catalogue records automatically from the Dublin Core metadata. For more info on this contact UVic

To stay informed about ETDs, sign up for the ETD-l listserv sponsored by the NDLTD.