

# **The Development of Open Access to ETDs in Canada: A Partnership between Canadian Universities and Library and Archives Canada**

**Sharon Reeves**

Manager, Theses Canada, Library and Archives Canada, Ottawa

*Keywords:* Theses Canada Portal, OAI-PMH, pilot project.

## **ABSTRACT**

A centrally coordinated program for theses accepted at Canadian universities was established in 1965 at the National Library of Canada (NLC)<sup>1</sup>. The program had two objectives: to acquire and preserve a comprehensive collection of Canadian theses and to make them accessible in Canada and around the world.

From 1965 to 2003 the program was delivered in a traditional fashion – paper theses were reproduced on microfilm or microfiche, which were then made available to researchers via interlibrary loan or purchase. Bibliographic records in NLC's catalogue made this often difficult to find research more easy to locate.

In January 2004 Theses Canada launched the Theses Canada Portal, which provides open access to over 45,000 electronic theses. In 2004 staff at Library and Archives Canada went on to develop its first OAI harvester and repository in order to harvest electronic theses and metadata directly from those universities with electronic theses submission programs.

In Canada at the present time there is strong and growing support for open access to scholarly publications. At the same time universities look to Theses Canada to take a leadership role in the transition from print to electronic theses. In this paper I will discuss the development of the OAI harvester and repository at Library and Archives Canada and the rapid pace at which Canadian universities are starting to embrace ETDs.

## **1. HISTORY OF THE NATIONAL THESIS PROGRAM IN CANADA**

The centrally coordinated national thesis program at Library and Archives Canada (LAC) was established in 1965 at the National Library of Canada (NLC) at the request of various deans of Canadian graduate schools. The program had two objectives:

- to facilitate access to theses by identifying what was available and by providing copies via ILL or sale; and,
- to preserve theses by microfilming paper copies and storing preservation masters according to archival standards.

For thirty seven years the program continued to fulfill its mandate in much the same way. The only significant changes to program delivery came first in 1990 when NLC outsourced the reproduction and sales aspects of the program and then in 1998 when, under the terms of a new contract, UMI Dissertations Publishing (UMI) began to digitize the theses as well as

---

<sup>1</sup> Now part of Library and Archives Canada.

reproducing them on microfiche. It is important to note that since the beginning of the program theses authors have signed a non-exclusive license with NLC, which initially allowed it to reproduce their theses on microfilm and which was revised in the early 1990s to allow for the distribution of theses in electronic format on the Internet.

## **2. PARTNERSHIPS**

Participation in the national thesis program was and continues to be voluntary. In 1965 five universities became the Canadian Theses Service's (CTS) initial partners. Over the past forty years participation in the program has steadily grown to the point where Theses Canada currently has partnerships with fifty-eight universities. One of the unique aspects of this arrangement is that there is no written agreement between each university and Theses Canada. The only criteria for joining the program is that a university must belong to the Association of Universities and Colleges of Canada. This makes for a flexible situation for universities, which can stop participating in the program at any time. It also leaves Theses Canada free to develop its program for its university partners as a whole, after undertaking any necessary consultations.

## **3. ETDS IN CANADA**

During the late 1990s interest in electronic theses and dissertations started to percolate at various Canadian universities in response to the initiatives concerning ETDS that were taking place in the United States and elsewhere. Several notable examples from this time period are Simon Fraser University, which experimented with using Adobe PDF format to mount its masters theses on the Web and the TUG (TriUniversity Group) Electronic Theses Project. Partners in TUG were the University of Guelph, the University of Waterloo and Wilfrid Laurier University. While this joint initiative ultimately failed, the University of Waterloo went on in 1999 to implement the first Canadian ETD submission program. Other universities investigating the possibility of ETD submission at the time were l'Université de Montréal and l'Université Laval.

## **4. NATIONAL CONSULTATION ON ETDS**

It was natural for the Canadian academic community to look to the National Library of Canada for leadership in those early days of transition from print to electronic theses and accordingly NLC organized a national consultation on e-theses, which took place December 4, 2000. Seventy-eight participants from across Canada took part in the consultation. Represented were deans of graduate schools, university administrators and professors, library professionals and students, as well as participants from various organizations associated with academia. The main objective of the consultation was to discuss the national theses program in an electronic environment. Out of the consultation came the proposal for a Canadian portal for electronic theses.

The second important issue discussed at the consultation was that of sales of Canadian theses by NLC's contractor. Concerns had been expressed by a number of graduate students and others to the effect that the agreement between NLC and UMI, allowing UMI to sell copies of theses, violated a fundamental principle to ensure freedom of access to scholarly information. To quote William Maes, University Librarian at Dalhousie University and a speaker at the consultation: "Does the current commercial situation lead us down a path that restricts access to the very things we produced? This issue is not that different from our current situation with journal publication where authors freely give up their rights in order to be published, while the

very reason for publishing is thwarted by the high cost of getting that information back. ... Part of today's discussion emanates from a belief that electronic theses will bring down costs and allow us to take back ownership of the distribution and preservation of our own works" (Maes 2000).

## **5. THE THESES CANADA PORTAL**

After the consultation a national advisory committee was established for the Canadian Theses Service. At the same time theses and systems staff at NLC got to work developing a proposal for a searchable collection of Canadian theses on AMICUS, Canada's national online catalogue. This ambitious proposal was finished by spring 2002. It included the addition of abstracts to bibliographic records, the development of a web-based search interface for the numerous theses and dissertations in NLC's collection and of several new search indexes specific to theses. It also included a proposal for the patriation of theses digitized for the national program by UMI. This proposal became the foundation for the Theses Canada Portal.

Over the course of the fall of 2002 and throughout 2003 the design and systems development for the Portal was done. On April 1, 2003, in keeping with the new direction of the national program and to match the name chosen for the Portal, the Canadian Theses Service changed its name to Theses Canada. Throughout 2003 UMI sent NLC electronic versions of all the theses it had digitized under the terms of the contract in effect between 1998 and August 2002. In all NLC received over 45,000 ETDs in PDF format from UMI.

Theses Canada catapulted into the 21<sup>st</sup> century with the launch of the Theses Canada Portal, which took place on January 29, 2004 at the Ontario Library Association Superconference. At the time of its launch it provided access to NLC's extensive collection of theses and dissertations on microform, a collection that currently numbers close to 250,000. It also provided free online access to the subset of ETDs received from UMI as well as extensive information about the national program. The Portal was further enhanced in 2004 by indexing the theses and dissertations to enable clients to search their full text. It can be accessed at <http://www.collectionscanada.ca/thesescanada>.

## **6. OAI AT LIBRARY AND ARCHIVES CANADA**

While the launch of the Theses Portal and the provision of open access to a large number of ETDs was a significant achievement, it still left the national thesis program dependent on sources outside NLC to realize its mandate. NLC's agreement with its contractor precluded the provision of immediate open access to the theses digitized on its behalf. What was needed was a way of, in the words of Bill Maes, taking back ownership of the distribution and preservation of Canadian theses and dissertations. The way forward was clear at NLC – the solution would have to be ETDs, anything else was unaffordable.

At its second meeting, in October 2002, the Canadian Theses Service Advisory Committee (now the Theses Canada Advisory Committee), struck a technical sub-committee composed of representatives from CTS, l'Université Laval and the University of Waterloo. The sub-committee was charged with resolving problems with respect to the transition to electronic theses in Canada and to develop the means for universities to send their ETDs directly to the NLC. This second objective was the direct result of concerns expressed at the national consultation in 2000. At the October meeting the Advisory Committee recommended that the Canadian Theses Service conduct a pilot project to acquire e-theses from l'Université Laval and

the University of Waterloo. Both universities had already established ETD submission programs.

The sub-committee began its discussions in January 2003 and its work progressed concurrently with the development of the Theses Canada Portal. Its first task was to establish how the NLC would acquire ETDs from universities. Right from the beginning the solution was evident. The NDLTD Union Catalog harvests metadata from its member institutions using the OAI Protocol for Metadata Harvesting. L'Université Laval had already set up its program to meet OAI (Open Archives Initiative) requirements and the University of Waterloo was taking steps to make its metadata compliant for harvesting. The advantages of setting up a national electronic thesis program at NLC based on OAI open source software were immediately apparent. Once NLC was set up as both an OAI service provider and data provider it could harvest metadata for ETDs from universities and in turn make it available for harvesting by other organizations such as the NDLTD. The metadata would also be converted to MARC 21 and uploaded to AMICUS, the National Library's online catalogue.

In developing its harvesting program NLC faced a challenge that most institutions do not have. In order to continue to expand its collection of theses and dissertations, thus fulfilling the institution's mandate to acquire and preserve Canada's bibliographic heritage for current and future generations, NLC wanted to harvest not only theses metadata but the electronic theses and dissertations themselves.

At the time that these early discussions took place the Government of Canada announced that the National Library of Canada and the National Archives of Canada would be combined to form a new institution – Library and Archives Canada. As of April 1, 2003 the two institutions would act as one, although the supporting legislation would not be passed for another year.

## **7. THE PILOT PROJECT**

In October 2003 the technical sub-committee met in Ottawa, along with representatives from LAC's Information Technology Branch and Standards Division, to make some decisions which were needed before the pilot project could begin. The sub-committee decided that LAC would harvest universities once a month. LAC would only harvest theses in PDF format and metadata would be in ETD-ms.

The sub-committee discussed the need for an identifier on the metadata records that would make it possible to organize the harvested theses in the directory on the server. At the meeting the idea of using ISBN numbers was considered but discarded as being too problematic. Instead staff at LAC developed a TC [Theses Canada] number which includes the ILL library symbol for each university so that the harvested theses could be organized by university on the LAC server.

The LAC harvester and repository were developed after the meeting and were ready by the end of 2003. An extension was written to the harvesting program to allow for the harvesting of the e-theses themselves. LAC had to make some changes to AMICUS in order to upload the harvested records and Waterloo and Laval had to make some changes to their processes as well. The target date for the start of the six-month pilot project was April 1, 2004. LAC staff intended to evaluate the pilot project in October 2004 in order to identify and resolve problems and to assess the resource implications at LAC.

In 2004 a number of factors conspired to slow the project down. One of the significant ones was the transformation to Library and Archives Canada, which resulted in the identification of a number of new initiatives. Many systems staff were deflected from their ongoing projects to work on major systems developments related to the new initiatives. In spite of this LAC started to harvest metadata and PDFs from Laval and Waterloo.

The delay meant that Theses Canada was able to bring two more universities into the pilot. The University of Saskatchewan started its ETD submission program in December 2002 shortly after the CTS Advisory Committee meeting. In July 2004 university personnel contacted Theses Canada to ask when it would be possible to send its metadata to LAC. Theses Canada, after consulting with the systems personnel responsible for the harvester, invited the university to join in the pilot project. The university had developed its ETD submission program using the ETD\_db freeware from Virginia Polytechnic Institute and State University, which allowed a thesis to include multiple PDF files. The harvesting program at LAC, on the other hand, had been written on the basis of harvesting only one file. Participation of the University of Saskatchewan allowed LAC to identify this weakness in its program, which will be fixed when sufficient time and resources are available.

The University of Manitoba began to participate in the pilot project in fall 2004. Its Faculty of Graduate Studies gave its approval to set up an ETD pilot project in May 2004, using the DSpace institutional repository software that the University Library had already started to implement. By November the university already had eight ETD records in its repository. As a result of some communications between systems staff at the university and LAC the records were harvested. This turned out to be extremely serendipitous in that it was discovered that DSpace only supports metadata in the Dublin Core metadata standard. As a result, the programmer at the University of Manitoba developed an DC to ETD-ms crosswalk, which the university plans to make available to other implementers of DSpace for ETDs.

The delay also gave the Manager of Theses Canada time to prepare two sets of guidelines during the summer of 2004. A draft document on harvesting requirements was prepared and sent to all stakeholders for review. At the same time guidelines on *How to Set Up an ETD Submission Program* were made developed. Since April 1, 2005 these guidelines have been available on the Theses Canada Portal at <http://www.collectionscanada.ca/thesescanada/s4-210.1-e.html>.

## **8. PILOT PROJECT EVALUATION**

The 693 harvested records from all four universities were finally converted to MARC 21 and uploaded to AMICUS on March 30, 2005, six months later than expected. In April 2005 LAC staff reviewed the records, which they found to be good quality access records although the records do not fully comply with international standards. However there were several problems with the conversion that needed to be addressed, as well as some issues concerning internal workflow to resolve.

Conversion problems were identified in the MARC tag 529 abstract note and, for French records, in MARC tag 655 – index term: genre/form. The problem with tag 655 was easy to fix by making a simple change to the conversion program. The problem in MARC tag 529 proved more difficult to resolve. In some records punctuation and mathematical and scientific symbols in the abstract note were not converting properly from the metadata records. Staff identified that at least part of the problem was caused by the fact that students were using different coding in their source documents and cutting and pasting their abstracts into the university thesis submission template, i.e. some types of coding were not converting properly.

Two workflow requirements were identified. LAC cataloguers needed to know what new records were uploaded to AMICUS each month in order to make any required changes to the records. And in order to avoid having to add a copy statement to each record, cataloguers requested that the location of the URL on records in AMICUS and AMICUS Web be moved to the top of the record display, which is where it appears in the records on the Theses Portal. This change will result in considerable resource savings for LAC's Resource Description Division.

The resolution of the various conversion problems made it possible to finalize the guidelines on Theses Canada's harvesting requirements. The harvesting program should be available to other Canadian universities sometime in the fall of 2005.

## **9. STATUS OF ETD INITIATIVES IN CANADA IN 2005**

In 2002 when the Canadian Theses Service started to develop the Theses Canada Portal, the University of Waterloo was the only Canadian university with an established ETD submission program. L'Université Laval had finished its pilot project and was just starting to accept ETDs. In three short years interest in ETD initiatives has burgeoned. Theses Canada conducted a short survey in 2004 which revealed that out of its fifty six university partners that did not have ETD submission programs twenty two, or 40%, were either interested in setting one up or were in the planning or pilot project phase. Interest has been confirmed by the number of calls Theses Canada receives from universities requesting information and assistance in setting up their programs. Replies were not received from 48% of Theses Canada's partners but, of these, 63% represent universities with very small graduate programs. Only five universities (9%) expressed no interest in ETDs. A follow-up survey of non-respondents will be done in fall 2005.

Since the survey was done a certain amount of shifting has taken place and at the present time at least three universities are ready for LAC to harvest their data repositories. Eleven more are involved in either planning or conducting pilot projects for ETD submission. These include some of Canada's large and prestigious universities such as McGill University and Queen's University. The DSpace institutional repository software is proving to be a popular choice – at least four universities have implemented it. The next survey will also follow-up with these universities and, at the same time, will encourage them to participate in LAC'S harvesting program.

## **10. CONCLUSION**

In the next two to three years many of the universities currently in the planning or pilot stages of their ETD initiatives will put ETD submission programs in place. This will be a busy period for Theses Canada as it starts to harvest theses and metadata from many of its partners while at the same time maintaining the traditional program it provides through its contractor. However in three years Theses Canada will be well on its way to taking back ownership of the distribution of Canadian graduate student research and to providing open access to it.

## **11. REFERENCES**

Maes, W. (2000). *Consultation on Electronic Theses*. Retrieved June 13, 2005, from <http://www.collectionscanada.ca/icbn/040021-401-e.html>.