

## A National Portal for Canadian Theses at the Library and Archives of Canada

Sharon Reeves

Library and Archives of Canada

sharon.reeves@nlc-bnc.ca

395 Wellington St., Ottawa, Ontario, Canada, K1A 0N4

www.nlc-bnc.ca

Keywords: portal, Canadian, theses

### Abstract

*Theses Canada has coordinated a centralized theses program for Canadian universities since 1965. Its traditional mission has been to acquire and preserve a comprehensive collection of Canadian theses and to provide access to them within Canada and throughout the world. As many Canadian universities move towards electronic theses submission programs this mission has expanded to include the acquisition and preservation of Canadian theses in this new format and to support Canadian universities during the transition period.*

*To this end a theses portal is being developed for the Library and Archives of Canada (LAC) website. The goal is to make the portal a comprehensive repository of freely available Canadian electronic theses and dissertations. Phase One of the Theses Canada online portal is expected to be launched by the end of 2003 and will include bibliographic records for the over 225,000 theses in the LAC collection, as well as approximately 45,000 full text theses digitized for Theses Canada by UMI Dissertations Publishing during the period 1998-2002. Planning for Phase Two of the portal development, which will permit Canadian universities to submit e-theses and metadata directly to the Library and Archives of Canada, is already underway.*

### History of Theses Canada

The Canadian Theses Service was created in 1965 as a program of the National Library of Canada, now the Library and Archives of Canada (LAC),<sup>1</sup> at the request of the Canadian academic community. Its purpose was to microfilm theses accepted by Canadian universities and to make them more widely available. Titles were catalogued for listing in *Canadiana*, Canada's national bibliography, and the microfilmed copies were made available directly from the National Library through inter-library loan or by purchase.

The program has gone through many changes since its inception 38 years ago. For many years the microfilming and sales were done in-house. Since the early 1990's the reproduction and sales of theses have been contracted out, most recently to UMI Dissertations Publishing.

Recently, to mark the transition from print to electronic theses at the LAC, the name of the Canadian Theses Service was changed to Theses Canada.

Participation in the Theses Canada program has always been voluntary. At the present time 55 universities participate in the program representing most major Canadian universities, including Université de Montréal and University of Toronto.

There are currently over 225,000 theses and dissertations in the Library and Archives of Canada collection. Titles are still listed in *Canadiana*, although now this is a CD ROM product, and are also included in UMI's database *Dissertation Abstracts International*. Catalogue records for Canadian theses and dissertations are available on AMICUS Web, <<http://www.nlc-bnc.ca>>, Canada's national online catalogue that is freely accessible world wide through the Internet.

### Theses Canada Mission

The traditional mission of Theses Canada has been to **acquire** and **preserve** a comprehensive collection of Canadian theses at the Library and Archives of Canada, to **provide access** to this valuable research within Canada and throughout the world and to **disseminate** the products by inter-library loan and direct purchase. But times are changing and this is no longer enough. Many Canadian universities are looking to Theses Canada to lead the transition from print to electronic theses.

This transition has resulted in an expansion of the Theses Canada mission to include the provision of support to Canadian graduate schools as they gradually make the transition to electronic theses and dissertations. Theses Canada is also developing the means for universities to submit electronic theses directly to the LAC. It commits itself to the acquisition and preservation of a freely accessible electronic theses collection through the Library and Archives of Canada.

<sup>1</sup> The Library and Archives of Canada is a new agency of the Government of Canada that combines the National Library of Canada and the National Archives of Canada.

## Requirements for a National Electronic Theses Program

A number of requirements for a national electronic theses program have been identified. These include:

- Develop and maintain comprehensive national coverage, i.e., include all universities and all masters' theses and doctoral dissertations;
- Maintain theses in digital form through their entire life cycle;
- Store theses in a stable digital format;
- Preserve digital versions;
- Enable theses to be full-text searchable;
- Ensure theses undergo a minimal number of conversions;
- Develop the capability for free public access to the corpus of Canadian theses from a single web-based access point;
- Adopt the ETD metadata standard to describe theses;
- Facilitate student supplied metadata;
- Guarantee student prescribed access rights; and,
- Ensure there is no redundancy in workflow between universities and Theses Canada.

In order to accomplish these ambitious goals Theses Canada has adopted a phased approach.

## Theses Canada Online Portal - Phase One

Theses Canada is taking the first step towards the creation of a national electronic theses program by creating a theses portal on the National Library of Canada website<sup>2</sup> <<http://www.nlc-bnc.ca>>. that will eventually function as a comprehensive repository of freely available full text Canadian electronic theses.

A proposal for Phase One of the Theses Canada online portal was prepared in the spring of 2002. Later the proposal was refined and funding was secured. By December 2002 development of portal functions and appearance began. The Theses Canada portal will be launched in the second half of 2003.

Initial content on the portal will consist of:

- Approximately 45,000 full text electronic theses processed from December 1997 until August 31, 2002 that have been digitized in PDF image format from hardcopy.
- Bibliographic records for all theses and dissertations in the Library and Archives of Canada collection. Records will include abstracts when available.

- Extensive information about the Theses Canada program including how it operates, who participates, information about copyright for students and much more.

Contrary to the LAC cataloguing policy that each format of a publication receives its own bibliographic record, it has been decided that records for theses will describe both the fiche and electronic versions. This is due to the enormous resource implications associated with creating separate records for such a large quantity of electronic theses.

Information about Theses Canada is already available on the Theses Canada web pages <<http://www.nlc-bnc.ca/thesescanada>>. The functionality and appearance of the web pages is preliminary at the moment. A fully operational design for the portal which will make the presentation of information more attractive and streamlined will be completed prior to its launch.

The search interface that is currently under development will permit searching of the entire database of theses or electronic theses only. It will be possible to do a basic search using the following indexes: any keyword, title, name, subject, abstract, AMICUS number or ISBN. It will also be possible to do an advanced search using Boolean operators for up to three index terms. Index terms for the advanced search will include all the same ones as the basic search plus degree and university and will provide three optional limiters - degree date, format and language.

## Phase One Challenges

There have been a number of challenges associated with the development of Phase One of the online portal.

The first was to secure funding for the project development. Theses Canada was successful in obtaining funding from the Digital Library of Canada Task Force (DLCTF) and from the Information Technology Services Branch (ITSB). The DLCTF is providing the funding for the portal design, staff to assist with the design of the site and the search interface, and to set up the directory structure for the electronic theses. ITSB is funding the hardware requirements and the staff needed to do the necessary programming.

There have also been some technical challenges associated with getting the electronic theses from UMI Dissertations Publishing.

Finally it has sometimes been difficult to coordinate the internal workflow between Theses Canada, the DLCTF and ITSB.

However there is sufficient time before the launch to resolve any outstanding problems.

<sup>2</sup> The websites of the National Library and the National Archives have not yet been combined.

## Theses Canada Online Portal - Phase Two

The launch of Phase One of the Theses Canada portal will provide access to bibliographic records for all the theses in the LAC collection, as well as to an extensive body of electronic theses, but it does not fulfill all the requirements of a national program for electronic theses. Phase Two of the portal will allow Canadian universities to submit electronic theses and theses metadata directly to the Library and Archives of Canada. The preliminary development of Phase Two is being funded by the LAC's Strategic Investment Fund.

While the primary focus at the LAC is currently on the development of Phase One of the online portal work has already started on Phase Two. An analysis of the current workflow for print theses has led to the identification of some scenarios for future workflows for electronic theses.

A basic operational workflow is expected to include the following elements:

1. Electronic theses will be sent to the Library and Archives of Canada from universities via FTP for inclusion in the online portal. It is likely that the LAC will first accept e-theses only in PDF text format but will build the capacity to accept them in XML at the earliest opportunity.
2. Theses Canada will become a service provider, harvesting metadata records for approved theses and dissertations from universities using the OAI protocol for metadata harvesting.
3. Theses metadata will be made available for harvesting by other organizations such as the NDLTD.
4. The harvested metadata records will be converted to the MARC format before the records are loaded into AMICUS at the LAC.

A number of tasks have to be accomplished before the Library and Archives of Canada can accept electronic theses directly from universities.

1. A student submission template is required in order to simplify the workflow at the LAC. This means that only one program will be required to convert ETD-ms records in the XML format to the MARC format before loading the records into AMICUS. Theses Canada, in consultation with the University of Waterloo, is currently developing a standard submission template for students based on the ETD-ms. The template will be easy for students to fill out. It will be available to universities on the portal when it is ready.
2. An ETD-ms to MARC crosswalk is required. It is currently under development at the LAC.
3. The capacity to harvest metadata records using the OAI protocol needs to be developed. The Information Technology Services Branch has already done some

work on this since a number of projects in the Library and Archives of Canada require it.

By the fourth quarter of the 2003/2004 fiscal year Theses Canada expects to be ready to conduct a pilot project to harvest metadata and acquire electronic theses directly from one or several universities. The University of Waterloo and Université Laval have been identified as possible participants since both universities already have well developed electronic theses submission programs. The pilot project will permit the identification of problems that require resolution before the option is made available to all Canadian universities.

### Phase Two Challenges

The number of challenges associated with the development of Phase One of the portal pale in comparison to the challenges of Phase Two.

First, development of the student submission template has proven more complicated than expected due to the need for it to be easy for students to fill out and generic enough that it can be used at any university.

Next, a number of changes will be required to the internal workflow for theses at the LAC, primarily in the cataloguing process.

Two parallel SYSTEMs will be in operation at the Library and Archives of Canada during the lengthy transition period from print to electronic theses. Most universities will continue to submit their print *and* electronic theses to Theses Canada's contracted service provider. At the same time various universities will gradually begin submitting electronic theses directly to the LAC.

Unless the records can be programmed to merge on loading this has the potential to result in duplicate bibliographic records for electronic theses, one originating from the UMI workflow and the other harvested from individual universities.

By far the biggest challenge facing Theses Canada in developing a program for direct submission of electronic theses to the Library and Archives is that of coordinating with so many universities at various stages of transition. Accommodating the many perspectives and ways and means to participate in the Theses Canada program will be an organizational challenge. Some universities are not at all ready to move to electronic theses submission programs. A few universities, such as Université Laval, are far ahead of other Canadian institutions and are moving rapidly towards mandatory submission of electronic theses from students. The Theses Canada program will continue to provide leadership and accommodate all of its participants.

## Digital Preservation

The issue of digital preservation still needs to be satisfactorily resolved, not simply for electronic theses but for all electronic publications. Many Canadian universities are reluctant to give up the security of having their theses reproduced on preservation quality microfiche. For this reason it will be necessary for Theses Canada to continue its reproduction service for the foreseeable future.

## Conclusion

The issue of how to accommodate the needs and requirements of individual universities while maintaining a strong, centralized national program will be dealt with university-by-university, and issue-by-issue. The goal remains to acquire, preserve and provide access to a comprehensive collection of Canadian theses and dissertations, either in electronic format or in fiche, at the Library and Archives of Canada. Theses Canada will continue to provide leadership to meet these goals.