Canadiana, the National Bibliography for Canada, in the Digital Age

Liz McKeen,
Director,
Resource Description Division,
Published Heritage Branch,
Library and Archives Canada

Meeting:	**162. Bibliography**
Simultaneous Interpretation:	English, Arabic, Chinese, French, German, Russian and Spanish

Introduction

Good morning. I am very happy to be here this morning, representing Library and Archives Canada (LAC) and discussing Canada's national bibliography Canadiana in the Digital Age.

What I would like to do this morning is to explore with you the changes that are taking place in the concept and the practice of national bibliography in this Digital age. I will focus my remarks on the experiences in Canada, where we are moving towards a redefinition of the scope and nature of the national bibliography, and we are increasingly recognizing the need, and seeking opportunities for, collaboration with others in creating the national bibliography.

Digital Publications at Library and Archives Canada (LAC): The Story So Far

Legal Deposit of e-Publications

In Canada, the legal deposit law was extended to cover electronic publications as of January 2007. This has been significant for us, but not completely new; the National Library of Canada, one of the predecessors of Library and Archives Canada, had been collecting electronic publications since 1994, and we had amassed a small but interesting collection now numbering some 30,000 titles, or 405 gigabytes of data. These titles are almost all "out-of-copyright" or crown copyright materials which we were able to make available to the world on our Website. They have each been duly catalogued using the Anglo-American Cataloguing Rules and other standards, and listed in the national bibliography.

While the law in theory includes Websites, databases, geospatial data and many other kinds of digital materials in its definition of publication, as a practical matter we have for now limited the application of legal deposit to quasi-traditional digital publications which take the form and outward appearance of print monographs and serials. The legal deposit law also requires that publishers send to LAC any metadata that is available describing their publication.

Web Harvesting

In 2005, LAC first began to experiment with the use of the Web crawler Heretrix to scoop up from the Internet large numbers of whole Websites, at least a snapshot of each Website as it existed on the day it was captured. We began with Websites of Canadian federal government departments, with an initial crawl in 2006, and we now have a regular schedule of crawling federal government websites twice a year. We have also done a crawl of the ministries of the ten Canadian provincial and three territorial governments, and are planning to do more of these on a regular basis as well.

The "Government of Canada Web Archive" went live, publicly available on the Internet in November 2007. We made a point of designing the interface in such a way as to make it very clear and evident to users that this was an "archived" version rather than a live version of the Websites; we did not want Canadian taxpayers using the archived version of the Canadian Revenue Agency website to pay their income taxes, for example.

The Government of Canada Web Archive contains a number of iterations or snapshots of about 1,500 websites. What is the content of these Websites? Our first crawl of the .gc.ca domain netted some 40 million digital objects (2 terabytes of data) which yielded some 900,000 PDF files. Each subsequent crawl has netted more and more digital objects. On our last crawl of federal government Websites, we captured over 70 million digital objects and 3 terabytes of data. Are these publications? Are they serial issues? Are they forms that citizens fill out to obtain government services? Are they monographs, duplicated every time we take another snapshot? Are they new publications?

Mass Digitization of the Collection

While LAC had been digitizing individual collection items, maps etc. for a few years, in 2007, LAC undertook a program of systematic mass digitization of its collections. This program is still in its early stages, working out an overall strategy while at the same time undertaking a number of digitization projects, many in collaboration with others, including the Canada Gazette (Canada's official government publishing instrument), Hansard (the verbatim record of Parliament), the Canada Yearbook (published by Statistics Canada), and a number of key census projects (of great interest to genealogists). Digitization challenges are many, and include funding and governance, selection of materials to digitize, collaboration with other digitizers, respecting copyright, and designing an efficient physical operation of a mass program, either in-house or out-sourced. A major requirement for digitization is that effective metadata be associated with all digitized items, so that users can find what they are looking for.

Canadiana in the Digital Age

Digital Publications at LAC

- e-Publications and selected Websites
  - Legal deposit of e-publications, January 2007
  - e-collection of 30,000 titles, 405 gigabytes
- Web harvesting
  - 1,500 Canadian federal department websites
  - Canadian provincial and territorial websites
- Mass Digitization of LAC collection

Digital Publishing

In summary, the mass of Canadian digital publishing is enormous and growing. This publishing is current, reflects the mentality and creativity of the nation, and is an important record of Canadian thought. Increasingly, digital publishing is supplanting more traditional formats. It cannot be ignored.

Canadiana in the Digital Age

Digital Publishing

- Enormous quantities
- Current
- Reflects & records the nation's thought
- May supplant traditional publishing formats over time

Cataloguing Digital Publications

Libraries' traditional, title by title cataloguing of publications is an expensive process, which is not easily scalable to handle the digital load.

In 2005, LAC held a "Description Day", attended by about 100 staff members from both the library and archival sides of the institution, to talk about innovation in description. One of the groups spontaneously formed at this event was a group bent on exploring how to "Catalogue the Web". This group quickly concluded that, because of the vast numbers involved, it could not be done in the traditional way. In 2006, LAC produced a cataloguing policy proposal, which we felt was both radical enough to encompass the enormous quantity of new digital publishing, while remaining steadfast on principles of access. This "Resource Description for Digital Publications" is available on the LAC Website.

Resource Description for Digital Publications: Policy

Briefly, in recognition of the massive numbers of digital titles, and the pragmatic need for partnership and collaboration in their description, this policy proposes an array of access solutions.

Canadiana in the Digital Age

Resource Description for Digital Publications: Policy

LAC policy on cataloguing of digital publications (2006)

Partially implemented 2008; future implementation depends on systems functionality

Available on LAC Website

1 <u>Basic</u> - Basic access to all digital publications to be provided through full-text indexing and searching. Ideally, this search capability should be made available to users through a public interface that features a high quality, sophisticated search engine able to rank results for users in a meaningful way.
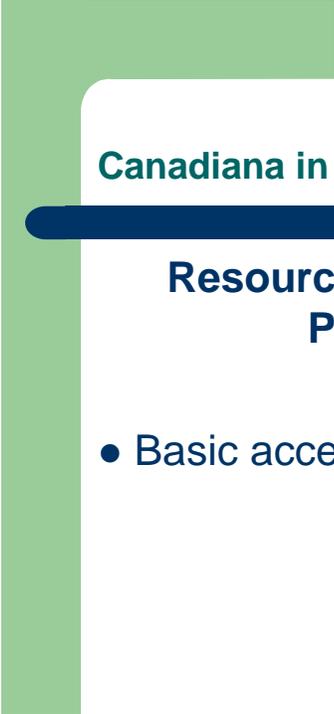
Canadiana in the Digital Age

Resource Description for Digital Publications: Policy

● Basic access : full-text search

2 <u>Supplementary</u> - Supplementary access through metadata to be provided. However, the creation of metadata to be done in the most cost-effective manner, using a continuum of possible means:

2.1 First choice: use metadata supplied by others : publishers, other libraries, other institutions, authors, students etc.. Descriptions would be used more or less "as is", as long as they can be made MARC-like.

2.2 Second choice: use metadata automatically generated or extracted from digital publications themselves. This form of artificial intelligence has not yet been fully explored; may require further study.

2.3 Third choice: use descriptions created by LAC staff for acquisitions purposes. These records are normally sufficient to identify an item, but do not usually include access enhancements such as subject headings, classification numbers, associated name authority records. Acquisition records may be created using metadata supplied by others or automatically generated.

2.4 Last choice: use international standards (MARC, AACR, ISBD)to create a full bibliographic description, with enhanced access such as bilingual subject headings, Dewey and LC classification, authoritative headings, all applicable notes, etc.

This choice is the most expensive option, and would be reserved for those few titles carefully selected from the mass of digital titles for special treatment. By providing a standard bibliographic description, LAC would be able to highlight the document, giving it greater prominence and accessibility.

Canadiana in the Digital Age

Resource Description for Digital Publications: Policy

- Supplementary access, in approximate order of preference:
  - Metadata supplied by others
  - Metadata automatically generated from digital object
  - Metadata created for other purposes (acquisitions)
  - Full, standard bibliographic records

When will we use the "full bibliographic record" option? Our proposed criteria for deciding which publications will get the full access treatment are:

- Those titles that document seminal or topical events, episodes, incidents, experiences in Canada (examples: elections, SARS crisis, natural disasters etc.).
- Titles intended for a special or priority collection within LAC, as defined by the Collection Development Policy.
- Titles that are intended for use as research or reference tools.
- Titles for which an LAC staff member has requested a full bibliographic record.
- Titles that fall under special bibliographic programs such as Cataloguing in Publication (CIP).

These criteria remain quite general, as they have not yet been refined through experience.

Canadiana in the Digital Age

Resource Description for Digital Publications: Policy

How will we select digital titles for full cataloguing?

1. Titles that document seminal or topical events, experiences in Canada (elections, SARS crisis)
2. Titles that are part of an LAC special or priority collection

Canadiana in the Digital Age

Resource Description for Digital Publications: Policy

How to select digital titles for full cataloguing?

3. Titles that are research or reference tools
4. Titles for which cataloguing is requested by LAC staff
5. Titles falling under special bibliographic programs (Cataloguing in Publication)

3. Digitized – Digitized titles have been drawn in most cases from the LAC collection for reproduction. In many cases, a good quality bibliographic record exists for the original item. This Resource

Description for Digital Publications Policy proposes that the digitized version simply be added to the existing record for the original. This single record approach is not consistent with the need for separate records for each format, as required by standard cataloguing rules, and with the usual requirements of a national bibliography. However, LAC sees this as a cost-effective approach that creates significant benefits for the end user in creating a simplified bibliographic record and a streamlined "one-stop" user search experience.

Canadiana in the Digital Age

Resource Description for Digital Publications: Policy

- Digitized titles – digitized versions of originals in the LAC collection will be added to the bibliographic record for the original
- Single record approach

Impact on the National Bibliography

National Bibliography:

The concept of Universal Bibliographic Control as articulated in the 1970s by IFLA hinges on the notion that each national library or national bibliographic agency take responsibility for listing all the publications of its own country in a national bibliography. With all countries working together, a comprehensive listing of the world's literature is possible, at least in theory. The consequences for world scholarship, the sharing of knowledge, the spread of ideas, the enrichment of cultures, and the advancement of humanity are enormous.

In addition, the benefits of a comprehensive national bibliography, retrospective and current, accrue to each country; a national bibliography is a window on the nation, providing a succinct yet representative reflection of the nation's culture, expression and thought over time. The ability of researchers or more casual seekers to discover and obtain access to the works of this country is greatly expanded through the existence of a current national bibliography.

Canadiana in the Digital Age

National Bibliography

- Universal Bibliographic Control
- Comprehensive listing of the world's literature, country by country
- National bibliography: a window on a nation
- Enhanced access through comprehensive listing, unambiguous identification of titles

In Canada, the current national bibliography began in 1953 with monthly printed issues of Canadiana, which over the years became more comprehensive in scope and coverage. At the same time, staff spent many hours compiling a retrospective national bibliography reaching back in time to 1755, when printing was introduced to Canada, more than a century before it became the country it is today. Today, Canadiana is embodied by the bibliographic records residing online in AMICUS, our Web-based bibliographic system, and in the form of an annual CD ROM edition of Canadiana which contains some 2 million records.

National Bibliography and Digital Publications:

When we look at the nature and sheer scale of digital publishing, we can see very quickly that the traditional concept of a comprehensive national bibliography listing all publications of a given country may not be possible or even desirable.

In terms of quantity, digital publishing vastly outnumbers and will continue to outnumber traditional publishing for many reasons. For example, Internet publishing is more available to more people with low cost, little risk, and a much reduced editorial infrastructure. This can result in more democratic publishing (e.g. wikis), greater variety in the quality of publications, more ephemera (e.g. blogs and email), and more and more varied authors and other creators.

In terms of nationality, it is difficult to assign a Website to a particular geo-political entity or nationality. Certainly this information is often not readily evident from the Website itself. Websites by their nature are equally available around the world.

Canadiana in the Digital Age

National Bibliography & Digital Publications

- Quantity of digital publications (electronic serials & monographs, Websites, email, wikis, blogs, Facebook, etc.)
- Nationality of digital publications unclear

Canadiana: Impact of the Resource Description for Digital Publications Policy

In Canada, LAC's implementation of the Resource Description for Digital Publications Policy has the potential to affect the national bibliography in a number of ways.

Comprehensiveness: *Canadiana* has aimed until now to be a comprehensive listing, both retrospective and current, of all Canadian imprints, as well as titles by Canadian authors publishing outside Canada, and a selection of titles about Canada or on subjects of great interest to Canada. Under the new policy, a relatively small subset of digital publications will be identified in the national bibliography per se, based on the selection criteria mentioned earlier. Other Canadian digital publications would not be listed. The establishment of authoritative name headings will also be less than comprehensive.

Unambiguous identification: *Canadiana* provides separate listings for each edition and format of every title. Under the new policy, digitized versions of titles will be added to the bibliographic record for the original, thus creating a record reflecting aspects of both publications.

Authoritativeness: While *Canadiana* can remain an authoritative reference source, it may not remain the single, authoritative source for all Canadian publications, as it would no longer be comprehensive and unambiguous.

Canadiana in the Digital Age

Resource Description for Digital Publications Policy: Implications

- **Less able to approach aim of comprehensiveness**
- **Unambiguous identification of titles (expressions, manifestations) not always possible**
- **Authoritativeness as an information source could be compromised**

Implications for *Canadiana*

LAC has charted this course for a number of reasons:

1) Balance – LAC recognizes the value of digital publications as vehicles of Canadian thought and creativity. The current cataloguing policy favours traditional publications over digital publications. This bias may not be justified, and will become increasingly less viable over time, as digital publications become increasingly important. With the Resource Description for Digital Publications Policy, LAC aims to achieve a better balance in coverage of Canadian titles, regardless of format.

2) Leveraging Technology – LAC is seeking to make use of technological aids to provide access to digital publications through full-text indexing, automated or semi-automated generation of descriptions, and emerging features of digital publications themselves. This approach is cost-effective and has the potential to provide very good access.

Canadiana in the Digital Age

Resource Description for Digital Publications Policy: Implications

- Will achieve better balance in coverage of traditional and digital publications
- Leverages technology to assist in providing access
- Affordable

Collaboration

The collaborative model has a long and glorious history among the libraries of the world. Traditionally LAC, like all libraries, makes maximum use of bibliographic records created and shared by other libraries whenever possible. LAC is a prime creator and sharer of source records for Canadian imprints and other Canadiana. LAC is also a supporter of the Program for Cooperative Cataloging coordinated by the Library of Congress to provide a framework for maximizing the shared creation of high-quality bibliographic and associated data, and of the ISSN network and other collaborative efforts.

Given the scope and scale of digital publishing in all its forms, LAC and many other libraries are looking to the collaborative model to achieve reasonable digital coverage. However, the collaborative model in the digital age becomes both more important and less organized. LAC is increasingly interested in finding ways to obtain and make use of metadata in many forms, created not just by libraries but by others operating in a different context, such as publishers, authors, students, book jobbers, individuals with many interests and areas of expertise etc.

<u>Metadata from Students</u>

<u>Theses and Dissertations</u>
We have been obtaining thesis abstracts written by students and adding them to bibliographic records for some years; the abstract is full-text searchable, and provides enhanced access by subject. In recent years we have used OAI (Open Archives Initiative) to harvest electronic theses from certain Canadian universities, along with the associated metadata produced by the student. This metadata is automatically massaged and added to AMICUS, our online bibliographic system, as a MARC record containing somewhat non-standard content.

<u>Publishers' Metadata</u>

<u>Trusted Digital Repository</u>
LAC is developing a Virtual Loading Dock, a component of a Trusted Digital Repository that will allow Canadian publishers of digital publications to deposit these titles with LAC, and to provide metadata on them via a Web form. The metadata will then be transferred to a metadata repository for use in providing access to these titles.

LAC is also looking to the TDR systems package to provide a technological means to automatically generate metadata from the digital object itself as it is being ingested. This functionality has not yet been developed.

<u>ISBN metadata</u>
LAC has developed an automated ISBN system which provides for the import and export of bibliographic data in the ONIX format, with the capacity to convert ONIX to MARC records for bibliographic listing. ONIX records will be transferred to LAC's Cataloguing in Publication (CIP) program, where cataloguers will use it as a base and add enhancements and standards.

Canadiana in the Digital Age

Collaboration:

Compiling the National Bibliography by Using Metadata Supplied by Others

- Students' metadata on their thesis or dissertation
- Publishers' metadata
  - When depositing legal deposit e-publications in Trusted Digital Repository (Virtual Loading Dock)
  - With ISBN request

Data from Individuals

LAC has experimented with several projects designed to capitalize on the knowledge of individual users. Data provided by individuals needs to be identified as such, as it is not authoritative in the institutional sense. While this metadata is highly non-standard, it has a unique value and appeal to readers that can add a very personal dimension to our understanding of collection materials.

Faces of War: In a project called Faces of War, Website viewers are asked to comment on digitized photographs and identify individuals in the photos if possible. This project can become very personal, as viewers identify their grandfather and his young friends from another era (for example). The comments are briefly vetted by a staff member, then made public in a way that makes clear that they represent personal commentary.

Canadiana in the Digital Age

Collaboration:

- Metadata from Individuals
  - Faces of War http://www.collectionscanada.gc.ca/faces-of-war/index-e.html
- Example of Comment on Photograph:

    My father Abe G….. is the second from the right. He arrived in france July 6 /1944. Dad married a british gal named Joan Mary W……
    They had 4 children , Frances, Gary , Wayne and myself
    Derek..Dad rarely talked of the war but I know it changed him in some ways . however if he had not been there , I would not be here. War has many strange twists.
    Dad passed away 1981.My deepest thanks for having this picure for me to find online . Thank you.
    Derek

Project Naming: LAC photographic collections depicting the people of Canada's North consist of thousands of photographs dating from the late 1800s to the mid-20th century. However, very few Inuit in these images were identified at the time these photographs were taken. The people of Nunavut (one of three Canadian territories located in the far north) have never had a chance to assist in identifying these individuals because the collections were located far from their communities and, prior to digitization, there was no means of easily transporting the photographs to Nunavut. LAC created a Website in 2001 containing digitized photographs and asked Inuit elders and youth to identify people portrayed in the photographs. The naming of these still anonymous people has become very time sensitive. Today's Elders may be the last people able to identify these individuals from the past, whose names might otherwise remain lost forever. This project is being conducted by LAC in collaboration with the Nunavut Sivuniksavut and the Nunavut Department of Culture, Languages, Elders and Youth, and involves youth visits to communities.

Nitrate Photographic Negatives (Project under Development): LAC has a significant collection of older nitrate negatives for which no metadata exists. The project under development is to digitize these historic photographs and to develop a Web interface to ask users to add identifying information to them. We will seek information from viewers on the people, places, events, and objects portrayed in the photographs. This metadata will be compiled automatically into a user-generated finding aid, with suitable caveats, which may be updated, corrected or amended by other users in a process similar to the Wikipedia.

Canadiana in the Digital Age

- Metadata from Individuals
  - Project Naming
    http://www.collectionscanada.gc.ca/inuit/index.html
  - Nitrate negatives (project under development to obtain user-generated metadata)

Conclusion

I have explored with you some of the changes that are taking place in the concept and the practice of a national bibliography in the digital age, through the Canadian experience. The experiences in Canada, while still new and tentative, are moving us towards a redefinition of the scope and nature of our national bibliography. We are increasingly recognizing the need, and seeking opportunities for, collaboration with others in creating the national bibliography. We want to collaborate in new ways, including exploiting the knowledge of individual Canadians, whether it be expert knowledge or personal, anecdotal contributions. Through our Resource Description for Digital Publications Policy, which we are implementing incrementally as opportunities allow, we are attempting to recognize and adapt to the realities and the opportunities of digital publishing and digital access. I have no doubt that in the months and years to come, these early experiences will be vastly broadened and deepened to a richer understanding of the possibilities and the obligations of a national bibliographic agency in the early years of this century.

Canadiana in the Digital Age

Conclusion

- Redefining scope & nature of national bibliography
- Seeking collaboration with new partners
- Seeking to obtain new types of metadata
- Adapt to realities & opportunities of digital

The published output of a nation can be considered to contain the distilled thoughts, attitudes, culture and creativity of the citizens of a country. It will remain an important role for national bibliographic agencies to provide a true and full reflection of this output, and to ensure that the world has access to the publications of the nation. While the end remains constant, the means by which this reflection is created and presented to the world will be a shared experience, shaped by digital opportunities in the future.

Canadiana in the Digital Age

Conclusion

- Publications are distilled thought of a nation
- Continuing role of national bibliography– to provide a true and full reflection of a nation's published output, and to ensure access to it for current and future generations