

The Connected World -- Opening Electronic Doors at the National Library of Canada

Lucy J. Bottomley, Nancy Brodie

National Library of Canada

e-mail: lucy.bottomley@justice.x400.gc.ca, nancy.brodie@nlc-bnc.ca

Abstract

For hundreds of years, libraries have been the custodians of the accumulated knowledge of human civilization - collecting it, preserving it, and making it available and accessible to those who need it or want to use it. This tradition, supported by a variety of legislation, is continuing. With the advent of technology, libraries have changed how they do their work but not why they do it. As we move into the 21st century, the National Library of Canada's basic mandate to collect, preserve and provide access to Canada's published heritage remains unchanged. What has changed over the years is the nature of the published heritage. It now includes electronic publications. Technological progress has expanded (and made more ambitious) our information needs, having placed at our disposal a greater access to national and international information resources. Consequently, the National Library of Canada has broadened its mandate. It not only provides access to the country's published heritage. It also endeavours to facilitate access for all Canadians to world-wide networks of information resources.

There are many complex challenges related to providing or facilitating an equitable and universal access to these networks. Creating and managing an electronic collection, preservation or archiving of this electronic collection, and investing in a technical infrastructure required to support a large national heritage collection are the three challenges that will be highlighted. The world becomes connected at an economic expense. Some of the costs are related to the lack of standards, others to the continued inoperability of systems. With digitization costs ranging from \$2 to \$6 per page and given the ever decreasing support for cultural agencies, other financial solutions have been sought. The National Library has become engaged in various (government and private industry) partnerships for its multiple digitization activities that include digitizing unique manuscript collections, creating electronic versions of its current and past exhibits, and making available in electronic format some unique bibliographic tools. These activities will be briefly reviewed.

World Wide Web, the final destination for the National Library's digital information, is gradually becoming the backbone of the National Library's services and operations and many resources are dedicated to making this happen. The Library's publishing program is now driven by technology and more than one format may be used when publishing electronically. Based on user requirements, some publications are retained in print or at

least are available on a print-on-demand basis. Some of the web-related issues that will be reviewed are: equitable access, including accessibility for persons with disabilities and bilingualism (English and French).

We are learning how to manage digital information and we are committed to making this information as accessible as possible to all Canadians - and to the world - as the nation becomes connected. The conclusion emphasizes the management aspect of National Library's World Wide Web, the digital library of today.

Introduction

For hundreds of years, libraries have been the custodians of the accumulated knowledge of human civilization - collecting it, preserving it, and making it available and accessible to those who need it or want to use it. This tradition, supported by various legislation, is continuing. With the advent of technology, libraries have changed *how* they do their work but not *why* they do it. As we move into the 21st century, the National Library of Canada's basic mandate to collect, preserve and provide access to Canada's published heritage remains unchanged. What has changed over the years is the nature of the published heritage - it now includes electronic publications. The National Library of Canada has defined the term "digital library" as the digital "face" of the traditional library, comprising electronic collections, access tools and services which complement rather than replace those of the traditional library. Technological progress has expanded (and made more ambitious) our information needs, having placed at our disposal a greater access to national and international information resources. Consequently, the National Library of Canada has broadened its mandate: it not only provides access to the country's published heritage, it also endeavours to facilitate access for all Canadians to world-wide networks of information resources.

There are many complex challenges related to providing or facilitating an equitable and universal access to these networks. Creating and managing an electronic collection, preservation/archiving of this electronic collection, and investing in a technical infrastructure required to support a large national heritage collection are the three challenges that will be highlighted.

Creating and managing an electronic collection

In the digital environment, libraries can build an electronic collection through creation of original electronic publications, digitization of retrospective materials and collection of externally published digital resources. This section concentrates on collection of external networked information resources.

Collection Policy

The National Library began its electronic collection by identifying and acquiring copies of 12 Canadian serials published on the Internet in 1994. As the volume of Canadian Internet resources grew, we developed a definition of a remote access electronic publication and selection criteria. Our sole means of collecting was by acquiring copies. We were firmly on the "ownership" side of the access versus ownership paradigm. Our selection criteria excluded "Web sites per se." We collect publications within Web sites. But as the National Library of Canada we recognized our role in providing access to Canadian information and

developed a service called “Canadian Information by Subject” which provides links to information about Canada from Internet resources around the world. The Canadian Information by Subject subject tree is based on the Dewey Decimal Classification and is available in English and French. With three years of experience we are now developing a Collection Management Policy which encompasses the above collecting activities. It recognizes that the decision to link to a remote site or document is a collection decision and needs a policy basis.

The current draft of the National Library Policy and Guidelines for Remote Access Electronic Publications recognizes three levels of collecting: Archived, Served, and Linked. These levels are adapted from the Berkeley Digital Library SunSITE Collection and Preservation Policy. At the Archived level, a publication is hosted or mounted on the National Library site and the National Library has made a commitment to keep the intellectual content of the publication available on a permanent basis. At the Served level, the publication is hosted or mounted on the National Library site, but the National Library has not made a commitment to keep it available permanently. This status can be applied as an interim status for a publication awaiting a decision as to Archived status or can be applied to a publication for which there is definitely no commitment to preserve and another organization is recognized as having primary responsibility for the publication content and its maintenance. At the Linked level, the publication is hosted elsewhere and National Library provides a link to the remote site. The National Library has no control over the publication but is committed to the on-going maintenance of the link. The National Library will concentrate its activities on publications at the Archived level in order to ensure permanent access to Canada’s electronic heritage.

Selection

Selection criteria must evolve as the electronic publishing world evolves. The current focus of National Library collecting is on Internet publications from the federal government of Canada. The easiest titles to identify and acquire are those that most closely resemble print publications. However, the National Library recognizes that the most important titles to be collected are those “original” electronic publications which are not published in any other form. It is difficult to recognize original electronic publications particularly on sites of publishers who also publish in print.

Standards are an important aspect of selection. The National Library allocates priority to acquiring publications in standard, non-proprietary formats. Whenever possible, we acquire and store electronic publications in the formats in which the publications were published. However, we recognize that it may be necessary to convert non-standard format publications to a standard format for preservation purposes.

Acquisition

For the Archived and Served collecting levels, the National Library acquires copies of networked electronic publications and stores them on its server. Such acquisition is carried out via e-mail, FTP, mirroring or downloading, depending on the nature of the publication and the agreements which can be reached with the publisher. At present the National Library requests permission before acquiring any publication. We do not use a spider to

automatically acquire publications. For resources to be collected at the linked level, the National Library maintains a link from specific areas of its Web site.

Intellectual Access

Ideal tools for intellectual access should support access by browsing, access through an integrated catalogue, access via an index or table of contents and a full text search access. The National Library provides two points of access to its Electronic Collection, from a special page on its Web site and via the library catalogue. The Web access supports browsing and includes a full text search. Both access tools provide direct access to the full text of the electronic publications. Index access is not currently supported.

Access to linked resources is available by subject and a title keyword search (Canadian Information by Subject). We are currently discussing the effectiveness of different access tools for linked resources.

Organization for Storage

At present the National Library organizes the publications in its electronic collection in a hierarchical directory structure. The key attribute governing the organization is degree of access. Certain publications are available to all Internet users; others are restricted to in-house use. At the next level of the hierarchy monographs and serials are distinguished. The organization of files for serial publications mirrors the organization of the serial check-in records: active or inactive, government publisher or commercial publisher. This organization is serving us well at the present but will be modified as we build a more sophisticated infrastructure for this collection.

Preservation/archiving the electronic collection

It is a basic policy of the National Library to preserve electronic publications electronically. We feel that linear media such as microfilm will not be able to adequately preserve the content of multifaceted networked electronic publications.

Archiving should preserve the integrity of an electronic publication and guarantee its authenticity. The Report of the Task Force on Archiving Digital Information describes preservation of information integrity as “to define and preserve those features of an information object that distinguish it as a whole and singular work”, and lists the features of integrity as “content, fixity, reference, provenance and context”. Publications must be complete and an archive should be comprehensive within its defined scope of collecting. There should be assurance of reliable, ongoing access to archived publications.

A strong case can be made for national libraries to assume a major responsibility for archiving electronic publications. In Canada, this role is an extension of the National Library’s traditional role. It is argued that archives of electronic publications are a public good. Archives are infrequently used and are unlikely to be a profitable endeavour for commercial firms. National libraries are in a position to support a public good; national deposit collections are authoritative sources of a country’s publications; national libraries take a long-term view, and are unlikely to “go out of business”. Under the principles of the International Federation of Library Associations and Institutions’ Universal Availability of

Publications (UAP) Core Programme, each country is responsible for providing access to its publications. National libraries are addressing the matter of electronic archiving in the context of their UAP role. In Canada, it makes sense for the National Library to have the infrastructure development and funding to support electronic-publications archiving.

The National Library has been archiving Canadian networked electronic publications since 1994, originally as part of its Electronic Publications Pilot Project (EPPP) and subsequently as a permanent electronic collection. How do the Library's policies and practices address the archiving goals described above? What are the key challenges?

Integrity

Current electronic publications unavailable in non-electronic format (original electronic publications) are a priority for acquisition. The National Library gives preference to electronic publications issued by the originator rather than taken from a secondary source. Early acquisition from the originator offers the best chance of preserving a publication's integrity. The Library endeavours to preserve the content and the "look and feel" of all archived electronic publications. The Library also acquires and stores versions of a publication that has been changed, and is developing standard terminology for describing versions. These descriptive labels must be obvious to those who reach a publication by using a search engine, so that the user cannot readily discern which version retrieved through a search is the most up-to-date. Whenever possible the Library acquires and stores electronic publications in the formats in which they were published.

Another challenge is preserving links. HTML is considered a good format for archiving because it is a non-proprietary standard accessible through many browsers. The EPPP recommended that the Library interpret the boundary of an electronic publication as normally being the linked objects on one Internet domain. This has been refined by adding "which are judged to be intrinsic to the publication." But acquiring, storing and providing access to even this limited set of links has proved challenging. What is necessary is for publishers to produce "transportable" publications, and for libraries to have better standards and tools for managing hypertext documents. The Library has not yet adopted "time-stamping" and "digital signatures", but considers them essential to preserving integrity.

Authenticity

The National Library's standard bibliographic records are viewed as authoritative sources, and the Library intends its records for its electronic collection to be equally authoritative. Works in the electronic collection are being catalogued for *Canadiana, the National Bibliography*.

Completeness

The National Library collects only complete electronic publications, and does not acquire abridgments, abstracts, tables of contents or promotional sites. Where several versions of an electronic publication exist in different formats, the Library acquires all standard, non-proprietary formats unless versions in proprietary formats are more complete. For instance, if a publication is available in ASCII and PDF (Portable Document Format), but certain tables are only available or readable in the PDF version, the Library will acquire the PDF version. The Library does not compress the electronic publications in its collection. Standard serials

check-in procedures are applied to electronic serials. Automated support for claiming electronic journals is still being investigated, and Web spiders or robots may help in this area.

Assurance of access

An archive must be accessible with current technology, preferably with widely available non-proprietary technology. The National Library's archive is accessible through the World Wide Web, but documents within the archive are in a variety of formats. Some formats are not accessible to text-based Web browsers, and some require proprietary software to view or print. The Library facilitates access to both standard and non-standard formats of publications.

Survivability

Electronic media are fragile in comparison to print or microform, and can be used only with specific hardware and software that may have lives even shorter than the electronic media themselves. The Library creates daily back-ups of its electronic collection, thus refreshing the digital information. However, 20 years of experience in preserving machine-readable cataloguing records has established the necessity of migrating digital information from one hardware/software configuration to another. Conversion of non-standard format electronic publications to standard formats may be required. It may be impossible to preserve a replica of the original electronic document during the process, so the objective will be to preserve the integrity of the document.

Technical Infrastructure

To support the National Library's mandate in this electronic age and to keep pace with the general trend in libraries to integrate digital collections, systems and services a sound technical infrastructure is de rigeur. The National Library has a modern bibliographic system, AMICUS, which it developed in partnership with CGI Inc. The Library has developed a sound telecommunications infrastructure for its Internet services. Pilot software based on Windows and UNIX file managers was developed to support management of the Electronic Collection. And various hardware and software has been acquired to support digitization and publishing activities. However, the existing infrastructure was not capable of supporting significant digital library activities on a large scale and a new enhanced infrastructure had to be planned for and created. This was not an easy task especially with library budgets continuing to shrink. Eventually, the Library opted for a three-phase Digital Library Infrastructure Plan:

Phase I - now complete - aimed at improving the end-user content creation functionality, specifically for digital projects and for the web content creation and management.

Phase II - still in progress - will establish such key components of the technical infrastructure as an extended relational database management system (RDBMS), a full text search component and an application development environment. At least one digital library application will be tested against those two components.

Phase III will implement additional components of the technical infrastructure, such as electronic document management system, enhanced web site management, and storage management systems. End-user development tools and additional server software will also be considered.

The completion of Phase II will allow the Library to proceed more efficiently and effectively to develop its digital collections and services, and to engage in more complex digital library

activities. The infrastructure, when fully implemented, should provide the Library with a solid and flexible technical foundation that can be enhanced as technology evolves.

Digitization initiatives

Digitization is one of three important methods of building electronic collections. The National Library has adopted this working definition of "digitization": it is the process of converting the content of physical media (e.g., periodical articles, books, manuscripts, cards, photographs, vinyl disks, etc.) to digital formats. At the National Library, digitization normally results in a document or resource that is accessible at the Library's Web site, and thus on the Internet.

The world becomes connected at an economic expense. Some of the costs are related to the lack of standards, others to the continued inoperability of systems. With digitization costs ranging from \$2 to \$6 per page and given the ever-decreasing support for cultural agencies, other financial solutions have been sought. The National Library has become engaged in various (government and private industry) partnerships for its multiple digitization activities which include digitizing unique manuscript collections, creating electronic versions of its current and past exhibits, and making available in electronic format some unique bibliographic tools. These activities are briefly reviewed.

Digitization is not an end in itself. It is based on what people need or want. Selection criteria, particularly those which reflect user needs, are of paramount importance. Therefore, the principles underlying traditional collection development still hold true when materials are being selected for digitization, but there are also many considerations related to technical, legal, policy, and resource considerations.

The National Library's digitization program continues to focus on selective digitization of reference tools and on thematic projects, with some limited digitization of portions of collections of special interest in Canadian studies. Essential to these ongoing efforts is the need to ensure that the Library builds on strengths, complements existing electronic and traditional collections, minimizes duplication of effort, and works in collaboration with other interested institutions and partners.

So why digitize? In the most basic terms, digitizing a document in print or other physical media (e.g., sound recordings) makes the document more useful as well as more accessible, and that means expanded service. Full-text searching of documents becomes possible, and hypertext links lead the reader to related items within the text itself as well as to external resources. Ultimately, digitization does not mean replacing traditional library collections and services; rather, digitization serves to enhance them." The National Library has completed multiple digital projects, some of which were made possible by the generous support of the private sector, with funding from Industry Canada's SchoolNet or through the general collection development process within the National Library. These projects can be accessed by going to the National Library's 'Digital Projects' address <URL: <http://www.nlc-bnc.ca/digiproj/edigiact.htm>>.

World Wide Web

The World Wide Web, the final destination for the National Library's digital information, is gradually becoming the backbone of the National Library's services and operations and many resources are dedicated to making this happen. The National Library's World Wide Web service, established in June 1995, is approaching its third anniversary. Because the Library is one of the primary resources for the support of Canadian studies, the Web is proving invaluable in making information on Canada's published heritage more widely available than ever before. The Web service also makes it possible for the Library to make its services available electronically, and to provide Internet access to other information resources, both in Canada and around the world. The Library's publishing program is now fully technology dependent and more than one format may be used when publishing electronically. Based on user requirements, some publications are retained in print or at least are available on a print-on-demand basis. Other formats are used and the decision is based on such criteria as document size, the expected use (on-line browsing, printing, downloading), and technical issues (bandwidth, user capabilities, availability of e.g. proprietary software). The Library, as an agency of the Government of Canada, is also obliged to comply with government-wide policies and legislation relevant to using the Internet as a communications tool within the federal government. The key ones, universal/equitable access, including accessibility for persons with disabilities, and bilingualism (English and French) are briefly reviewed.

Universal/equitable access

Universal accessibility ensures that sites are developed to serve the widest possible audience and the broadest possible range of hardware and software platforms, and that the needs of users with disabilities are considered. Equitable access policy within the Government of Canada institutions ensures that reasonable measures are taken to communicate effectively with disabled persons in a truly accessible electronic government. Since the change in web browsers and other web tools is rapid and continuous it is important to be aware of the developments and analyze the tools' enhanced functionality in terms of accessibility. Quite often the new features of these tools are inaccessible to a wide range of users, including persons with disabilities. The assistive devices rarely keep pace with the mass market web tools. The onus is on web developers to deliver the information in the way that allows everybody to benefit. At the National Library of Canada we ensure that all documents are developed with accessibility in mind, with text alternatives provided to all graphics, advanced features minimized and at least one version of our publications widely accessible. Some of the biggest improvements in accessibility will come from full implementations of HTML 4.0 and CSS2, and the National Library is looking forward to major browser and authoring tool manufacturers' implementations of these accessibility improvements. As we move into a universally connected world, where information is more and more enabled via the Web, it is critical to provide technology that enables all Web users.

Bilingualism

All communications posted on the WWW must be posted in both official languages (English and French) at the same time, ensuring the linguistic quality of the texts, and the presence of all required French language diacritics. This ruling reflects the obligations of Canadian federal institutions under the legal framework of the Official Languages Act and the Charter of Rights and Freedoms. Both ensure a respect and equality of status for the two official

languages of Canada. National Library's WWW service conforms to these regulations by making all documents available in both languages. Electronic links between the two language versions permit movement and flexibility. For publishers/web developers this means double publishing, from both the document preparation (editing, translation, formats, coding/programming) and document management (current storage, quality control, update, archiving) perspective.

Conclusion

There is a life cycle for networked electronic information that encompasses creation, organization for access and dissemination by publishers or document creators. The library component of the life cycle includes selection, acquisition, identification and organization for more precise access integrated with other published material, discovery and retrieval by clients and preservation for long term use. Through the wide range of its activities in the connected world, the National Library of Canada is involved in the total life cycle from creation through use.

Information technology has presented libraries with new and often complex challenges. National Library's World Wide Web, the digital library of today is just one example. At the National Library of Canada we continue our efforts to make a significant contribution to the development of electronic access to information in Canada to meet the challenges of the 21st century. We are learning how to manage digital information and we are committed to making this information as accessible as possible to all Canadians - and to the world - as the nation becomes connected. We make this commitment not just to Canadians today, but to posterity.

Bibliography

- Bottomley, Lucy, "Working on the World Wide Web", *National Library News*, 29:2 (February 1997). <URL: <http://www.nlc-bnc.ca/pubs/nl-news/enlnews.htm#1997>>
- Digital Library SunSITE Collection and Preservation Policy. Berkeley Digital Library SunSITE. 1996 <URL: <http://sunsite.berkeley.edu/Admin/collection.html> >
- Government of Canada. *Internet Guide*, 1997. <URL: <http://canada.gc.ca/programs/guide/> >
- Preserving Digital Information; Report of the Task Force on Archiving Digital Information*. Commission on Preservation and Access and the Research Libraries Group. 1996
- Starling Access Services. *Accessible Web Page Design*. <URL: <http://www.starlingweb.com/acc/actoc.htm> >
- Winston, Iris, "Building an E-Collection", *National Library News*, 29: 3/4, (March/April 1997). <URL: <http://www.nlc-bnc.ca/pubs/nl-news/1997/2903-e.pdf> >