

2018 IFLA International News Media Conference

“When Risk Becomes Real, Preserving News Becomes Critical”

18-20 April 2018

Old News is Good News: moving towards new approaches to newspaper acquisition and preservation at Library and Archives Canada

Annie Wolfe, Acquisition Librarian, Library and Archives Canada

Introduction

In 2004, the National Library of Canada and the National Archives of Canada merged and today Library and Archives Canada (LAC) is the proud steward of the foremost collection of Canadian documentary heritage in the world. With its annual budget of \$114,000,000 ¹ and over 900 staff, LAC serves clients in Canada and internationally and is dedicated to fulfilling its mandate which is:

- to preserve the documentary heritage of Canada for the benefit of present and future generations;
- to be a source of enduring knowledge accessible to all, contributing to the cultural, social and economic advancement of Canada as a free and democratic society;
- to facilitate in Canada co-operation among communities involved in the acquisition, preservation and diffusion of knowledge;
- to serve as the continuing memory of the Government of Canada and its institutions.

LAC’s collection is vast: 20 million items of published material; 250 linear kilometers of government and private textual records; 3 million architectural drawings, plans and maps; 5 billion megabytes of information in electronic format; 30 million photographic images; 90,000 films; 550,000 hours of audio and video recordings; 425,000 works of art; 550,000 music related items; and, last but not least, 2,300 newspaper titles. Newspapers reflect the unique realities of Canadian communities and are essential research sources for social historians, genealogists and others; however, they do present a number of acquisition, preservation and access challenges. This paper will outline LAC’s newspaper acquisition practices and the publishing trends driving change; describe current initiatives supporting development of LAC’s newspaper collections; and discuss a recent pilot on acquisition of digital news content.

¹ Source: Library and Archives Canada - 2016–17 Departmental Results Report, <http://www.bac-lac.gc.ca/eng/about-us/departmental-results-reports/2016-17-departmental-results-report/Pages/2016-17-departmental-results-report.aspx>

LAC's Newspaper Collection

The newspaper collection is one of the most heavily used collections at LAC. It is exclusively Canadian and includes over 2,300 titles. In addition to major daily newspapers, LAC holds a variety of community, regional, multi-cultural and Indigenous newspapers, as well as a small collection of historical print newspapers, 200,000 reels of newspapers on microfilm and 110 titles that have been digitized.

Newspaper acquisition: traditional mechanisms and opportunities to evolve

LAC's primary acquisition mechanism for published content is via the Legal Deposit Program. Like many other national libraries, LAC's Legal Deposit Program is enshrined in [legislation](#) and is governed by a suite of regulations. These regulations have evolved over time and have been regularly reviewed and expanded to include online publications, music, maps and other content. Canadian published heritage not subject to Legal Deposit may be purchased or received as a gift.

Newspapers in paper format were subject to Legal Deposit until 2009, when the organization decided to stop systematically acquiring print newspapers, in part because of the significant resources required to process, preserve and provide access to them. From that point, LAC focused on receiving newspapers on microfilm, as that format was more robust, and could be sent on inter-library loan. In addition, newspapers published on microform *are* subject to Legal Deposit: if between 4 and 99 reels are produced, the publisher must send one copy to LAC, and if 100 or more reels are produced, the publisher must send two copies to LAC. However, the landscape has continued to shift, and collecting and providing access to newspapers on microfilm is no longer sufficient.

Publishing trends driving new opportunities

Unsurprisingly, significant changes in the publishing industry, and in particular, the newspaper publishing model, are presenting LAC and other memory institutions with numerous opportunities to re-evaluate and refine their traditional collecting practices. In Canada, as in other countries, commercial production of newspaper microfilm has slowed in recent years and there are far fewer active producers operating. Those that remain have reduced their production runs so considerably that LAC is rarely able to consistently acquire them through Legal Deposit, which means the alternative is purchase.

As of April 2015², there were 105 daily newspapers and over 1,000 community newspapers in Canada but, as in other publishing industries, the news industry in Canada continues to consolidate and change. On November 27, 2017, TorStar and Postmedia – the two major owners of community newspapers in Canada – announced a sales transaction affecting 43 newspapers, which included the closure of 36. This trend toward closures is of great concern for many communities.

² Source: News Media Canada - Industry Info - FAQ, <https://nmc-mic.ca/about-newspapers/faq/>

Reduced availability of microfilm and fewer community newspapers are but two of the dramatic industry shifts driving institutions like LAC to seek solutions to the specific challenges regarding effective acquisition and preservation of Canadian news content. LAC has responded to industry trends by taking the opportunity to examine existing acquisition criteria and practices, and by continuing development of a robust digital infrastructure to support acquisition, maintenance and preservation of newspapers and other publications.

Deciding *what* news content to acquire is a layered consideration. There is a plethora of news delivery platforms generating content, new apps are emerging and even traditional concepts of an “edition”, a “daily” or a “journalist” have changed. Content included in a print edition may differ from a digital edition; many large newspapers feature dynamic news feeds on their websites that change throughout the day; paid subscriptions may include different content; and similar versions of the same newspaper may be branded differently. An institution like LAC must also balance the need to acquire national dailies while still ensuring the unique voices reflected in community, student and multi-cultural and indigenous newspapers are appropriately added to Canada’s national collection.

The sheer volume of digital news content may render comprehensive acquisition impractical and acquisition criteria based on the number of print subscriptions sold have become dated. Volume of digital content is also emerging as a key consideration in the context of Legal Deposit regulations as we ponder issues related to proprietary platforms, news apps and duplication of content in various file formats. Well documented selection decisions and vigilance regarding maintaining accurate holdings information is essential for responsible management of both analogue and digital news collections to prevent asymmetrical collecting practices and gaps in the collection. Developing and maintaining a collection of digital news content requires dedicated and significant technological assets, storage capacity and expertise.

Current work

These industry drivers have provided LAC the opportunity to rethink its newspaper collecting practices and to develop new strategies for the future.

In late 2016, LAC’s Chief Operating Officer established a cross-institutional working group to examine current interpretations of Canada’s Legal Deposit regulations and to compare LAC’s current acquisition approaches with those of other national libraries. The working group concluded that current Legal Deposit regulations provided sufficient flexibility for the near term but recommended some collecting practices, including those guiding newspaper acquisition, could be further clarified through the development of a new collection development policy, specifically for published heritage, with guidance provided in a suite of lower-level instruments.

This new Collection Development Policy for Published Heritage has been approved and LAC is now embarking on an innovative path of seeking to develop the lower-level instruments in direct collaboration with Canadians. This approach to directly engaging citizens in policy development is a new avenue for LAC and it will be interesting to analyse findings related to the level of participation, the effectiveness of this collaborative approach and the operational

implications of evergreen policy development. We know there is a high level of interest in our newspaper collections in particular and we look forward to hearing the ideas and opinions of our clients.

LAC has also contacted TorStar and Postmedia to begin discussion on how best to preserve the news archives of those community newspapers identified for closure. Given the high level of interest regarding news content, LAC is now working in partnership with other cultural memory institutions and the Ontario Library Association to develop a strategic proposal that will address the needs of all stakeholders. The first step will be undertaking a gap analysis study this Spring to better understand existing newspaper collections held by public libraries, how they are used, and how this group of dedicated partners might work best together to meet the needs of Canadians. This partnership may be a first step towards restoring a collaborative program of newspaper acquisition and preservation. Responsible management of news content on a national level requires contributions from a variety of partners.

These trends in the newspaper publishing industry are informing LAC's internal initiatives and are providing new lenses through which to assess evolving Legal Deposit and digital acquisition strategies. Swift and ongoing industry changes are prompting LAC to assess its newspaper collecting strategies and to seek opportunities to share this responsibility with other partners. By understanding the inherent complexity of news content and developing the appropriate expertise, we will ensure that essential Canadian news content is acquired, managed and preserved and, most significantly, that it is discoverable by all those that seek it.

Digital News Content Acquisition Pilot

Following a review of the literature, the published acquisitions team launched a Digital News Content Acquisitions Pilot over 4 months in 2017-2018 in order to better understand the practical and operational aspects of digital news acquisition, as well as to inform policy decisions. Many other national libraries have chosen to acquire PDF print masters as the most seamless and efficient manner to acquire news content. Obtaining digital news content will allow LAC to build a more comprehensive collection, eliminate the preservation burdens associated with print collections and reduce costs associated with purchasing microfilm. The focus of this small pilot was digital acquisition; specific analysis of preservation and access related to digital news content was considered out of scope.

The goals of the Digital News Content Acquisition pilot project were to explore the feasibility of obtaining PDF files directly from newspaper publishers; research and document realistic acquisition options for digital news, building on best practices and successful newspaper initiatives at other national libraries; and deepen internal acquisitions expertise regarding the ingest of digital news content and inform a framework for the development of the newspaper collection in the future.

Approach

The approach to achieve the goals of the pilot was divided into four stages, which were to:

- Validate our assumptions about PDF files by comparing format and delivery mechanisms of Canadian news publishing platforms and undertaking content analysis of PDF files, “behind paywall” digital editions, print editions and news websites;
- Test three digital acquisition options: manual ingest processes, existing web archiving capabilities and file transfer protocols;
- Liaise appropriately with publishers and LAC colleagues including Communications, Preservation and Reference Services to ensure a co-ordinated approach; and
- Make recommendations to inform next steps.

Selection of pilot project participants

Because the pilot project aimed to test three distinct digital acquisition options, selecting appropriate participants was essential. In order to test manual acquisition we identified five titles published by Northern News Services Online. These titles were selected because they are smaller weeklies produced in remote areas of Canada’s north, not available on microfilm and contain significant coverage of Indigenous communities, a current area of emphasis for LAC. To explore possibilities presented by existing web archiving and file transfer protocols, two medium sized daily newspapers were identified based on the following criteria: that the publication have both a print and a fee-based digital edition, that they be significant publications in their region, and that they reflect Canada’s two official languages. The first two newspapers we approached enthusiastically agreed to participate in the pilot project: *Winnipeg Free Press* (from the province of Manitoba) and *L’Acadie Nouvelle* (from the province of New Brunswick).

Manual acquisition of digital news

Understanding the complexity and level of effort associated with manually acquiring digital content from the website of a given publishers provides a useful baseline for comparison. We began this process by securing permission from Northern News Services Online (<http://www.nnsl.com/>) and testing a variety of internal processes. The process was indeed straightforward and was a simple matter of accessing the publisher’s log website, selecting and downloading individual files and transferring them to LAC. The following titles were successfully

acquired using this process: *Nunavut News*, *Northwest Territories News/North*, *Kivalliq News*, *Inuvik Drum*, and *Yellowknifer/Yellowknifer Weekend*.

While acquiring news content by manually downloading from the website proved straightforward and effective, it is only realistic when seeking a limited number of titles. The manual nature of this approach would quickly become impractical and increase the volume of work if we were to expand our scope to include more titles. While useful to quickly obtain at-risk content or to perform limited content samplings based on selected themes, manually acquiring individual files is not sufficiently effective to support LAC's longer-term digital newspaper acquisition goals.

Web archiving tools

LAC operates a successful web-archiving program, known for its high volume ingest capacity and preservation of full websites. The extent to which LAC's existing web archiving tools and processes could be used to ingest individual PDF files had not been fully explored. We collaborated with appropriate colleagues to evaluate what opportunities might exist for acquiring specific digital news content using the available tools *Adobe Acrobat Pro* and *Archive-It*.

Access to the digital news content was easily arranged. *L'Acadie Nouvelle* provided high-resolution and lower-resolution PDF versions of its daily newspaper via Google Drive. The *Winnipeg Free Press* provided login access to its e-editions portal, which grants subscribers access to the digital replicas of its print newspapers.

Adobe Acrobat Pro was explored using content provided by *L'Acadie Nouvelle* and it proved to be a useful tool. The reporting features of *Adobe Acrobat Pro* clearly illustrated the difference between the high-resolution and lower-resolution files and we were able to easily compare the content coverage of various versions. After reviewing the reports, it was clear the high-resolution concatenated PDF version was preferable because one file is easier to manage than multiples, high resolution is preferred for long-term preservation and, significant from an acquisitions perspective, it includes the same content and layout and the print edition in one PDF file.

For the *Winnipeg Free Press*, several web-archiving tests on the e-edition subscriber website were undertaken using LAC's subscription *Archive-It* service. The complex nature of the target websites presented significant challenges, most importantly, while effective for acquiring and preserving full websites, it was not possible to isolate individual concatenated PDF files for ingest.

From this work, we concluded that while Web archiving via *Archive-It* is effective when seeking to acquire and preserve full websites, it is not an effective means of acquiring selected PDF files only. Analysing and organizing the ingested content was difficult, resource intensive and high volumes of ancillary files were received. The scope of this pilot was limited to PDF files but as LAC prepares to acquire and preserve full news websites more systematically there are certainly

ways these tools and web archiving more broadly can advance LAC's news content collecting goals.

FTP – how far can you push (or pull) content?

While manually selecting individual PDFs files prior to transfer proved inefficient, FTP itself presented no significant technical concerns and supports bulk ingest as well. What warranted further consideration were the impacts different FTP based processes may have on LAC's services to our publishing clients. It is essential that contributing to the national collection be easy for publishers, and considerate of their business models and publishing workflows. Having publishers proactively "push" their content to LAC or requesting they provide access to enable LAC to seamlessly "pull" in their content are the two submission models requiring further exploration. Given the complexity and potentially high volume of digital news content to be ingested it is also essential that LAC be able to support and maintain new digital ingest services. Opportunities presented by FTP were evaluated though both a client services lens and from an administrative and operational support perspective.

Push refers to a client-initiated transfer of content to a LAC server. LAC currently receives some individual publications other than newspapers via a FTP push but nothing on the scale required by digital news content. Administratively, the push method is easy for LAC to set up and maintain. By providing the newspaper publishers with a FTP server address and a password the publisher can transfer PDFs directly to the LAC server. It is also possible to set up an internal trigger so that as soon as a new file is transferred, alerts are received by appropriate acquisitions staff and processing and data quality workflows can begin. However, LAC must be prepared to receive all content transferred by the contributing publisher and as such, the volume could be high in some cases. From a client service perspective, the push model is not entirely passive and responsibility to proactively transfer content remains with the publishers. However, most newspapers should be able to independently program transfers using tools they already have and LAC is well positioned to provide guidance as part of its ongoing outreach efforts to publishers.

Pull refers to an LAC-initiated transfer from the newspaper's servers. Administratively, the pull model requires LAC's acquisitions team to create an FTP account with each publisher that provides LAC with access to a specific folder on their server. When publishers have files to deposit, they proactively transfer them to this server where they can be retrieved by LAC. The pull method still requires active participation and a similar level of effort from the publishers' perspective as the push method. However, this model adds administrative complexity for LAC. Every time a newspaper publisher changes their password, upgrades their system, replaces a server etc. LAC would need to contact the publishers again to get the new information and reformat our automated systems. This will generate a lot of administrative work.

While we concluded that both push and pull FTP have merit to support digital news ingest, they vary on the level of effort required for both publisher and LAC. In each case, the publisher has responsibility for transferring content and must undertake some effort to initially establish new transfer processes, but it seems that neither model introduces significant submission burdens

from a client service perspective. However, since maintaining high-quality and efficient services to publishers is essential the push model presents less administrative burden for LAC and is the preferred model.

Overall, this small pilot project allowed LAC to test a variety of ways to acquire digital news content using currently available tools. It confirmed the utility of the concatenated PDF file, and a preference for using a push-based FTP ingest protocol.

LAC's ongoing digitization of newspapers

While the focus of this pilot project was the ingest of current, born-digital news content, there are initiatives underway in Canada to digitize analogue newspapers. In 2010, LAC established a productive partnership with *Canadiana.org*, recently merged with the Canadian Research Knowledge Network (CRKN), to digitize 109 copyright-free newspaper titles. LAC also facilitates Canada's National Heritage Digitization Strategy (NHDS) which is a pan-Canadian, multi-institution partnership committed to focusing the collective efforts of memory institutions in the area of digitization for discovery, access and preservation. The NHDS recently undertook a pilot project to digitize selected Indigenous newspapers. The first set of titles from this pilot project have now been digitized and are available online. In total, almost 600 issues of *Windspeaker*, published by the Aboriginal Multi-Media Society, were digitized and made available and another 1000 issues from two other publications will be made available soon.

Building the Infrastructure: LAC's Digital Asset Management System (DAMS)

Given the increasing amount of documentary heritage – including newspapers - in digital form, the appropriate technical infrastructure is paramount. LAC has approved a Digital Strategy to guide the renewal of legacy systems and introduce a robust digital asset management system. Digital ingest has been identified as a priority and the possibilities for digital news content are significant.

The DAMS digital ingest will provide LAC with capability to effectively transfer and package digital assets. In regards to published digital content LAC will work with selected publishers to test and deploy a system for:

- Improved efficiency for the transfer of digital assets to LAC through legal deposit;
- Enabling automatic linking of metadata and digital objects;
- Creation of a preservation-ready digital asset package from publisher provided digital assets.

The DAMS includes ingest tools that can accommodate both pull and push transfer models. These multi function file transfer features are sophisticated, secure and support the transfer of publisher-generated metadata. Before beginning to acquire the PDF newspapers with DAMS LAC will develop a process to acquire standard metadata from each newspaper publishers.

Voilà! Discovering newspaper collections

While it is critically important that news content is acquired and preserved, it must also be discoverable and accessible. LAC has recently entered into a contractual agreement with OCLC, an international non-profit library co-operative, for a new library system. *Voila*, the new OCLC-driven union catalogue supports discovery of holdings located in hundreds of libraries across Canada. OCLC system functions will allow Canadian and international researchers to discover Canadian newspapers and identify where they are located at libraries across the country, as well as provide a mechanism to understand coverage of newspaper content from sea to sea to sea.

Conclusion

LAC, its partners and its clients understand the important role newspapers and digital news content play in Canadian communities and around the world. Now, more than ever, it is essential that memory institutions like LAC continue working with partners to share expertise and be willing collaborators to ensure news content, generated by both existing and emerging content generators, be acquired and preserved. Continual shifts in the news media landscape are prompting fundamental conversations about truth, the value of an informed democracy and freedom of the press. Essential questions like this inspire us to build on the great work of others and continue to champion newspapers and digital news content as critical components of Canada's National Collection of published heritage.