

Digitizing newspapers:

A pilot project to support the National Heritage Digitization Strategy

Interim report – October 2017

Introduction

Newspapers are a rich, popular resource for serious and casual researchers alike, and many countries are undertaking largescale digitization projects to increase the number of newspapers available online. However, digitizing newspapers can be challenging given complex intellectual property considerations; fragile conditions; multiple formats; and fragmented collections.

In winter 2017, with the generous funding of the Salamander Foundation, the National Heritage Digitization Strategy (NHDS) embarked on a project to learn more about newspaper digitization in Canada and to add digitized titles to the national collection. This document describes the project and its implications.

About the National Heritage Digitization Strategy

The National Heritage Digitization Strategy is an effort to coordinate the digitization activities of Canadian memory institutions. Its goals are to focus the digitization efforts of Canadian memory institutions; identify standards and methods based on best practices; share knowledge and experience among partners; and protect documentary heritage most at risk.

An independent Steering Committee was established in the fall of 2016 to set direction, identify projects and develop funding models. In the planning phases, the committee identified the need for a pilot project to gather information on newspaper digitization in Canada. This project is the result.

Project Design

The project was designed to unfold in two phases:

Phase 1 – Consultation with organizations across the country with experience in newspaper digitization to gather best practices and recommendations.

Phase 2 – Digitization of a set of newspaper titles from Library and Archives Canada's (LAC) collection to test the lessons learned from the consultations and to add to the national discussion.



Citizens reading newspaper headlines concerning [Newfoundland's] confederation with Canada (Credit: Chris Lund / NFB / Library and Archives Canada / PA-128007)

Phase 1 – Consultations

The NHDS Secretariat consulted with seven representatives of six organizations and gained insight into: approaches, extent of collections, copyright, privacy, optical character recognition and formats.

The organizations consulted included:

- Bibliothèque et Archives nationales du Québec (BAnQ)
- Canadiana
- Library and Archives Canada
- Our Digital World (INK)
- Simon Fraser University
- University of Alberta

The key themes of the consultations were:

- The **number** of Canadian newspaper titles is vast and easy to underestimate.
- Organizations adopt **project-based approaches** as opportunities (funding, equipment, resources, and partnerships) present themselves. This can lead to inventive methods but fragmented results.
- Because of its relative stability and consistency, **microfilm** is by far the most common format digitized. Digitizing paper titles is mostly avoided because of their irregular formats and high cost.
- **Optical character recognition (OCR)** is imperfect but critical. Organizations invest ongoing resources to improve results.
- Clearing **copyright** can be daunting but is necessary. In many cases, publishers are open to discussing options.
- **Privacy** concerns are few, but need to be taken seriously. Organizations need to balance access to information with protection of personal information.
- Digitized newspapers appeal to wide **audiences** who value quantity over quality. Organizations almost universally adopt free and open access models.
- Organizations see **opportunities** to share experiences and build on expertise; aggregate collections for improved discovery; explore and support smaller community titles, particularly those using non-Latin scripts; and build digitization capacity in smaller organizations.

Phase 2 – Digitization

Drawing on its knowledge of its newspaper collection, LAC identified three possible areas of potential digitization and discussed these with the NHDS content strategy working group. Together the team chose to focus on digitizing First Nation's titles, because it most closely aligned with the principles of

diversity and inclusion found in the overall NHDS and the accompanying content strategy being developing by the working group.

LAC obtained permission to digitize from the publishers. The exact number of issues and the dates to be digitized will depend on how much can be done within the funds provided, but could include:

- [Windspeaker](#) – (2005 to 2015)
- [Turtle Island News](#) – (2001 to 2013)
- [Ha-Shilth-Sa](#) – (1974 to 2002)

Next steps

LAC staff continues to digitize the collection using the Salamander Foundation funds. LAC is also exploring options for optical character recognition and providing access to the digitized objects.

A report will be drafted outlining the results of the consultations and the digitization pilot. This report will be shared with the NHDS Steering Committee members and the broader Canadian cultural heritage community. The report can help inform how newspapers could be included under the NHDS. It will also share information on best practices with organizations across the country who are considering or actively pursuing newspaper digitization projects.

