

Sydney Morning Herald

21 May 2009

Newspapers online a 'fast track' to history

<http://www.smh.com.au/news/technology/web/newspapers-online-a-fast-track-to-history/2009/05/21/1242498844009.html>



The National Library of Australia aims to collate 4 million regional and metropolitan newspaper pages into a single searchable internet portal by 2011.

Photo: *Fairfax Photos*

Louisa Hearn

May 21, 2009 - 10:08AM

The Hawkesbury river boatmen who ferried wheat into Sydney in the early 1800s hold a unique place in Australian history.

Complaints of their "vexacious conduct" which caused untold damage to grain stocks leads the front page of Australia's first newspaper, and is now among millions of articles from our pioneering past available at the click of a button.

The National Library of Australia is the driver of this ambitious project, which aims to collate 4 million regional and metropolitan newspaper pages into a single searchable internet portal by 2011.

With 3.5 million news stories already live in the system, historians are hailing the [Australian Newspapers beta](#) search service as a fast track into our heritage.

"Newspapers were not inaccessible before but often you would have to travel from state to state for newspapers from capital city or rural areas," said Richard Waterhouse, professor of Australian History at the University of Sydney.

"This is not only going to make sources more accessible to academic historians but also for family historians who will be able to flesh out the story of their own family by finding references in local newspapers."

The process of transferring news from old-world library microfilm into digital format was no simple undertaking, said Cathy Pilgrim, NLA's director of Australian Newspapers Digitisation.

Firstly, the microfilm had to be scanned and sent overseas where optical character recognition software converted the image to text - essential for enabling keyword searches, she said.

With the NLA's core focus on material that is out of copyright, newspaper print quality can vary dramatically, often yielding less-than accurate results from character recognition software.

To complete the editing process, Ms Pilgrim said the NLA had leaned heavily on public goodwill, with users of its digital newspaper service already correcting more than 2 million lines of text in more than 100,000 articles and adding tens of thousands of tags.

"The human eye is much better than a machine and some of our correctors have been doing an amazing amount of work. Some find the work addictive or compulsive, and they are enhancing and enriching the service for future users," she said.

The public is also proving integral to search and rescue campaigns to find newspaper issues missing from the national archives. The NLA makes regular appeals to the public to check under their floor boards or sift through their attics for issues of [newspapers missing](#) from collections in libraries around the country.

Ms Pilgrim said that a user of the beta search service recently found reference in the *Brisbane Courier* to a newspaper called the *Braidwood Liberal* that was not even previously known to exist.

As it continues to grapple with the challenge of capturing the past, the NLA is also looking into the future in a separate project called Pandora, which seeks to capture a permanent record of significant digital publications and websites that are no longer publicly available.

"More publishing is being done online and less in traditional hard copy so the National Library really wants to be preserving online publishing as well. It's important to be storing and maintaining it so we can provide access to it in the future.

"The Sydney 2000 Olympics website was decommissioned a few weeks after the Games and is no longer available publicly, but through Pandora we managed to archive it," she said.