

Libraries reflecting communities: digitising NSW newspapers
Jerelynn Brown

Presented at the NSW Public Libraries Conference, *A fortunate life*
Mudgee, NSW, 11-14 November 2014

Abstract

Digitising newspapers has always been on the agenda for NSW public libraries and the State Library of NSW. 295 of 2,064 known NSW newspaper titles have now been digitised and are freely accessible via Trove. This is a story of collaboration at local, regional and state levels; putting the wisdom of the crowd to work, the changing ways clients can do research and the uses people have for digitised newspapers.

Introduction

Since enactment of the *NSW Copyright Act* in 1879, the State Library of NSW has had a statutory role in collecting and preserving all New South Wales publications through the Legal Deposit provisions of the *Act*. The Library collects new newspapers published in NSW extensively, and identifies and fills gaps in the collection with the object of a comprehensive collection. Newspapers reflect and document life in NSW and provide a reservoir of information to our clients. They are the most heavily used paper format at the Library as they are filled with articles of interest to researchers in all disciplines, at all ages and across the state and more widely.

The State Library has made a sustained strategic effort over many years, and has contributed to a number of local, state and national projects to acquire, provide access to and preserve NSW newspapers. As the NSW Convenor of the Australian Newspaper Plan, known as ANPlan, I will report to you about progress, and particularly, I would like to provide an update about newspaper digitisation, an area of increasing interest for our shared client group.

The newspaper collection at the State Library of New South Wales

The Library holds 2,064 newspaper titles published in New South Wales. 327 of these are current titles, that is, still being published. In round numbers, this includes more than 1,200 tabloids and 700 broadsheets with approximately 100 titles held in microfilm only.

The Library continues to acquire NSW newspapers, both newly published and those new to our collection. Despite news in the media about newspapers being 'dead', in 2013-14, the Library received eleven new NSW newspapers first published in that year through Legal Deposit. An additional 15 titles were acquired in 2013-14 that while not new, were new to the Library's collection!

Preserving newspapers

Microfilming provides access to recent issues and, critically, captures the page images for conversion to digital formats¹. The Library systematically microfilms all current NSW newspapers as every volume is acquired. The newspapers are then relocated to the climate controlled repository at Moorebank. The Library spends around \$400,000 annually to microfilm newspapers with more than \$300,000 spent on the current newspapers, leaving

¹ The NLA's Newspaper Digitisation Program takes a mass digitisation approach, and it is cost effective to scan the images from microfilm. This is cheaper than scanning from paper, particularly if the microfilm has already been produced, as an automated process with a higher speed may be used.

\$80,000 – 100,000 for retrospective projects such as unfiled sequences or newspapers that belong to another library and are loaned to the State Library for filming. Only a very small number of retrospective paper issues remain to be microfilmed.

Newspapers published in NSW are central to the State Library's collection and with the heavy use they receive by clients, it is important for the Library to continue to improve the conditions in which newspapers are kept and the access provided to them. Key initiatives include:

- Microfilming in three generations so that a pristine preservation master can be preserved; a copying master can be stored separately and retrieved only for scanning and a service copy is available for client use in the reading room.
- Relocating preservation masters for closed titles to a cold store, where they will be safely stored for several centuries.
- Reviewing NSW newspapers held at the other state libraries and the National Library which may not be held in the State Library of NSW's collection. The National Library recently repatriated issues from 183 NSW titles, and these will be used to fill gaps in the State Library's collection. There will be a positive impact on the completeness of the titles as missing issues are incorporated.
- Participating with the National and State Libraries in a 'search and rescue' strategy for newspapers known to be missing from the Library's collection. All newspapers through to the letter 'S' have been reviewed, and your support in 'finding' these would be appreciated. The Search and Rescue site is at <http://www.nla.gov.au/anplan/about/collect.html>

Why Microfilm?

When we handle the older newspapers printed on rag paper and compare them to later papers published from the mid-1800s when wood pulp became the basis for newsprint it is easy to see why there is such an urgent need to preserve the information in newspapers for clients of the future. Newsprint is an ephemeral carrier, perfect for the printed daily newspaper and reflecting the old adage, 'Newspaper today, chip wrapper tomorrow'.

Microfilming is still the preservation method of choice. Longevity is estimated for good quality polyester film, such as that used by the Library since about 1985, at about 500 years. However, pilots to capture born digital newspapers and prepress versions of print newspapers are underway under the auspices of ANPlan; and we expect that new digital preservation methods will become the norm in the next five years.

Finding newspapers

All newspapers known to be published in NSW have records on the Library's catalogue, and as well on the National Bibliographic Database accessible through TROVE. This adds value by maintaining an inventory of New South Wales newspaper publishing activity, and provides information about New South Wales newspapers to libraries and members of the public. It also contributes to world-scale discovery through Trove, WorldCat and Google. One of the Library's current projects is enhancing discovery for newspapers by providing separate records for online versions with links to online content to make it easier for clients to find what they require.

We encourage our clients to contribute. We are interested in their stories and newspaper histories and are looking for ways to add these to our records. One way to contribute is to tag and comment on the Library's catalogue records by using 'Library Thing', a cataloguing and

social networking site linked to the Library's catalogue and visible in the top right hand corner as you view records.

Another option is to contribute to a Wikipedia article. There are now over 213 of these articles live; and we are working to create at least a basic 'stub' article for every digitised title. This valuable initiative provides an opportunity for knowledgeable people, especially local history experts, to document the publishing history of each newspaper, and its relationship to its local community. Significantly this initiative takes content from our libraries to people in the place they are looking for it.

Each time someone does a Google search and finds a Wikipedia article about one of the NSW digitised titles, our content brings the client back to libraries – in the case of the newspapers, the link takes them to Trove, and the citations in the articles bring people back to the State Library. When local libraries put their citations into Wikipedia articles as they enhance them, people will be able to link back to their websites too. Library staff across the state have been active in adding a plethora of locally known content to these records, for example, the Orange *Leader* at [http://en.wikipedia.org/wiki/The_Leader_\(Orange,_NSW\)](http://en.wikipedia.org/wiki/The_Leader_(Orange,_NSW)); or Leeton's *Irrigation Record* http://en.wikipedia.org/wiki/Irrigation_Record.

Digitising newspapers

The Library is now into year three of our ten year program to digitise substantial portions of the collection. The Digital Excellence Program is funded by the NSW State Government and supports the opportunity to take the Library's key collections to people and enable access 24/7 in their place of choice. The program aims to make significant strands of collection material widely accessible – material of high interest and demand by clients, of potentially high use, treasures of the Library, and items of particular significance to the people of NSW. More than 20 million pages online will be generated for our clients through the program.

The program will also bring the Library's infrastructure up to date to enable more efficient working, and, to empower us to collect and preserve more born digital material, reflecting the transition in publishing from print to digital. Digitisation and digital collecting is quickly becoming business as usual.

Through the Digital Excellence Program the Library is digitising a range of selected content in addition to NSW newspapers, for example, the *NSW Government Gazette*; first fleet journals; Leichhardt's diaries; David Scott Mitchell's book collection; information documenting Indigenous peoples, the European discovery and exploration of Australia, waves of immigration and more. Many of you will have seen the Library's [World War I website](#) including diaries, maps, pictorial material, posters and ephemera.

Digitising NSW newspapers has always been on the agenda for the State Library and public libraries in NSW, and we are now getting traction as a result of the Digital Excellence Program. One of the Program's big success stories is the digitisation of NSW newspapers with more than of 5 million pages live on Trove. There are now 295 NSW newspapers digitised. In 2013-4 there were more than 16.7 million page views of NSW newspapers on Trove. This is as a result of local, state and national library collaboration.

The State Library of NSW was an early adopter of digital technologies to support the provision of information to clients and contributed to a number of innovative collaborative

projects. For example, the Ferguson Project² was the Library's first step in 1996 into the digital newspaper and magazine world.

Our work as a key partner with the National Library of Australia and other state libraries from 2004 resulted in *Newspapers Australia*. The National Library funded the infrastructure costs of this project and developed staff knowledge and expertise, processes, standards, and the Trove newspaper interface to search the digitised newspapers. The Program scanned microfilm, converted images of text into characters that are fully searchable through use of Optical Character Recognition (OCR) software and made this accessible along with facilities to search newspapers and support text correction.

The Library had been working collaboratively with the National Library to digitise newspapers for several years. Each year, with advice from clients and especially public library colleagues, the State Library identified priority NSW newspapers to be digitised. We provided a request to the NLA of our top 10 titles; and the NLA would fund the digitisation of some or all of them using the State Library's extensive collection of microfilm copying masters.

During this period we focussed on digitising newspapers from the Colonial period and those expected to have the widest interest, for example, the *Sydney gazette and NSW advertiser*, Australia's first newspaper. We participated in a number of funding bids to enable us to digitise more newspapers more quickly, but had no success. We supported several public library bids for grants and a number of these were successful with newspaper titles going live as a result.

When in 2011 the Digital Excellence Program funding was announced by the NSW government, a total of 59 NSW newspapers had been digitised and were live on Trove. We were getting great feedback, our clients loved it, but it was going to take 200 years to digitise NSW newspapers at this rate! Some public libraries had obtained grants or local funding and already had titles live on Trove. For example, Liverpool City Library had funded the digitisation of the *Liverpool Herald 1898-1907* right at the start.

The Digital Excellence Program became an important driver for our NSW newspaper digitisation strategy. After negotiating a Memorandum of Understanding with the National Library the program enabled us to add 1.5 million pages per year for four years. So with around 60 million pages of newspapers and 6 million pages funded, how did we decide where to start?

With the World War I commemoration looming and aware of what this would mean to local communities across the state, we decided to start with WWI newspapers. We have been using fairly tight selection criteria:

- newspapers published during the WWI period
- they must exist on microfilm of suitable quality and completeness
- must be of high interest or high use

²The Ferguson project began in 1996 as part of the Australian Cooperative Digitisation Project, a collaborative project between the University of Sydney Library, the State Library of NSW, the National Library of Australia and Monash University Library, with the support of the Australian Research Council. Serial publications in the date range 1840 to 1845 were selected from John Alexander Ferguson's *Bibliography of Australia* and include local newspapers, shipping and trade news, religious, colonial and political journals.

- often identified through advice from public library staff and suggestions from clients.
- newspapers of historical significance are preferred
- a range of geographical coverage is the aim
- newspapers to be digitised must be pre-1955 - that is, without copyright restrictions.

We started with newspapers that were published only during WWI, and then we did those with a small margin of years on either side of 1914-1918, and now we are working on those where WWI is only a dot in their publishing history, for example a current title like the *Cootamundra Herald* published from 1877 has been digitised and is accessible from the beginning up to 1954.

All newspaper titles we digitise are being digitised in their entirety up to 1954 in as much as we have copies and have microfilmed them. Digital technology will enable us to add missing issues later, so our current digitisation work also highlights the issues *not* held in the State Library or NSW public libraries at the time we microfilmed them.

The whole project currently relies on microfilm being available as this is an efficient and cost effective way to digitise the material. In fact, some of the newspapers we are scanning no longer exist in paper.

We are deliberately selecting newspapers from a range of towns and centres across NSW to ensure the best possible coverage of the state in accordance with our commitment to the NSW government to support regional NSW. In addition this gives people the experience of using a local newspaper on Trove – in many areas this has spurred local interest in raising funds to digitise more newspapers. For example, in the Central West zone of the state, out of 39 newspapers digitised at the end of the financial year, local funding or grants paid for 17 or 44% of all titles. The funding came from councils, historical societies, local companies and individuals. In 2013-14, 22 newspapers were digitised and went live with contributor funding - a great story of collaboration:

At the Library we encourage and support NSW public libraries in bidding for or raising funds to digitise local newspapers. In these cases if the digitisation is through the Newspaper Digitisation Program, the State Library will contribute by lending its microfilm copying masters for the scanning process.

The Library's capacity to digitise around six million pages or ten percent of known NSW newspapers through the Digital Excellence Program has given tremendous support resulting in nearly 15% of titles in the collection live and accessible freely on the internet. But the bottom line is that to achieve the aim of making all newspapers published in NSW and out of copyright available to clients online in a digitised version, we have a long way to go; and the magnitude of the newspaper digitisation effort in NSW needs support from all of us!

Progress to date

There are now 295 New South Wales titles live and accessible through Trove. At 30 June 2014 there were more than 4.7 million pages. State Library microfilm copying masters were used for all of the projects. New issues of digitised titles continue to go live daily and clients may register for an RSS feed on each title page to find out when new issues of a specific title are added.

Fourteen new titles funded by the Digital Excellence Program went live in the first week of October and another 22 titles, making a total of 36 new titles this financial year, will be live by May 2015.

In addition, we have given approval to contributors including public libraries, historical societies and a private individual to do another twenty titles this year through the contributor funding model, and we expect to see all of them live by June 2015.

Australia-wide, more than 15 million pages from more than 700 newspapers are now freely accessible via the internet. This includes newspapers from each state and territory and from the earliest published newspaper in 1803 through to the mid-20th century.

So how can someone find out what is in the pipeline for digitisation? Do a search for “[new titles coming](#)” on Trove to identify titles that are in the processing queue. This page currently lists newspapers to be digitised through Trove in 2014-15, whether funded by the contributor model, the State Library’s Digital Excellence Program, the National Library or other state libraries.

In an exciting development this year the first and second newspapers in languages other than English to be published in NSW, the French newspaper *Le Courier Australian* and *Guang yi hua bao* (The Chinese Australian herald) are on the list. The State Library is digitising these titles in partnership with NLA as they are pilots for the optical character recognition software. They will join two other NSW titles in languages other than English: in Italian, *Il Giornale Italiano* and in Estonian, *Meie Kodu* (Our home).

How do digitised newspapers work and what can our clients do to help?

When digitising newspapers the first thing is to scan the content, including creating web image files and assigning metadata. This enables the text to be read on a computer screen or mobile device. In effect this is just like reading the text on a good microfilm reader, but with benefits. In the case of newspapers, this is not very helpful because to find a specific article one needs to know the title, the date, the page or other identifying information to be able to find what you are looking for, in other words, you need to search.

So to make the text searchable the Optical Character Recognition software sometimes referred to as OCR, is run over the (high resolution) images, to interpret the marks on the page, to create, in effect, a transcript that is searchable. This involves importing OCR text, images and metadata into a software program.

This process is only as good as the print is uniform – with today’s print produced electronically, OCR software does a great job, but our early newspapers had common problems like uneven inking, non-uniform paper surfaces, and in some cases, there was only one set of type, so various type sizes and typefaces were used in the same article.

However, the good news is, the OCR product can be improved by using human intervention. This is what is meant by ‘Text Correction’. Text correction enhances the product, resulting in much more efficient and effective searches!

Text correction is absorbing for Volunteers – the National Library quotes around two million lines of text correction per month – They reckon if they paid the people to do the text correction, tagging and making lists it would cost something like \$12 million a year!

One great example is volunteers from [Penrith City Library](#). They correct text from the [Nepean Times](#) to make it more useful and accessible to family historians. This is a case where the volunteers had been indexing the newspaper, a very laborious process, until Penrith Library funded the digitisation and it became available on Trove. It is my understanding that a volunteer has been focusing on the war years and spent all of 2013 working on pages from 1914. (and is now working on 1915). Other volunteers have their favourite areas – one works on obituaries and family notices, and another is working on Council meetings for the years where they don't have the original minutes.

Ways researchers using the digitised newspapers

Increasingly APIs are being developed to access and explore the mass of newspaper data rather than to look for individual pieces of information. For example, [QueryPic](#), is an API that was developed to explore digitised newspapers through Trove. It enables researchers to enter a query and see during what time frame or geographic region the words are found in newspapers. This provides a new way of seeing, searching and understanding the digitised newspapers as a group, showing trends over time and mapping patterns.

Like NSW public libraries, the State Library has had an overwhelming response to newspaper digitisation from clients and communities. The technology we are using is enabling people to change the way they do research, changing their research skills, and their capacity and efficiency at research. Anecdotal accounts indicate that researchers who have spent years chained to large newspaper volumes in library reading rooms are now able to perform an increasing part of their research in coffee shops with good wifi!

To get a sense of the impact the project is having on our clients and people across the state here are some examples of what communities, special interest groups and researchers in the virtual world are doing with the newspapers besides text correction:

- The ANZAC biscuit recipes inspire many during the ANZAC day bakeoffs each year
- Fashionistas investigate fashion trends to find out about the evolution of designs, how designs travelled from Europe to Australia and even the use of today's endangered species for furry accessories
- Ravelry members find heritage knitting and crocheting patterns in our newspapers; socks are a favourite
- Real estate and property agents find history for heritage buildings and properties
- People on the land look at information about planting, yields, heirloom crops, pests and the weather
- Researchers on mining track commodity prices and mentions of old mining claims
- Meteorologists and environmentalists look for weather patterns and research climate change
- And no one could walk by digitised newspapers without mentioning genealogy – perhaps the biggest and most personal use by our clients.

We would be delighted to hear of other ways the community is using our digitised newspapers!

As newspapers are so heavily used by our clients and give so many different kinds of information to so many different clients, digitising NSW newspapers is a critically important part of our service to the community. The State Library's Digital Excellence Program has already, in just over two years, taken our collaborative effort to digitise NSW's heritage newspapers further than we could have hoped, enabling us to provide our shared client group

with a rich diet now and for many years to come, helping us to realise the vision of having NSW newspapers digitised and out there 24/7 in their place of choice for our clients.

On top of that, Wikipedia gives us a useful model of collaboration to build intellectual capital in the social media environment; and text correcting is a great opportunity for people to contribute the ‘wisdom of the crowd’.

It is our aim at the State Library of NSW to continue to develop the collection of newspapers, to make newspapers increasingly accessible to researchers online and to sustain an environment where researchers are empowered to find, share and create new knowledge. We look forward to working with public library colleagues to make this happen.

Appendix of practical recommendations

It is recommended that

- Newspapers identified as unknown NSW titles, including newly published titles in print and digital formats are notified to the State Library by public libraries
- Local government areas consider collaborating with community members to raise funds to digitise additional local NSW newspaper titles