A success story – Australian Newspapers Digitisation Program

Rose Holley[*]

This article gives an overview of the Australian Newspapers Digitisation Program (ANDP) and the resulting free online public service – Australian Newspapers service. The service was released to users by the National Library of Australia in August 2008 and has now been in operation for over a year. As at the end of October 2009, the ANDP has achieved free online public access to over 720,000 newspaper pages containing over 6.8 million articles. The public have helped to enhance the data by correcting over 6 million lines of text and adding 166,000 tags and 4,200 comments. The ANDP is collaborative, with every State and Territory library in Australia being involved. The service supports the key objectives of the Australian Newspaper Plan (ANPlan) to enable communities to explore their rich newspaper heritage. By mid-2011 the ANDP will have digitised and made available 4 million Australian newspaper pages dating from 1803-1954.

INTRODUCTION

The National Library of Australia (NLA), through its Australian Newspapers Digitisation Program (ANDP, http://www.nla.gov.au/ndp) and in collaboration with Australian State and Territory libraries, has made available an exciting new service that provides free online access to selected out-of-copyright Australian newspapers. The Australian Newspapers service (http://www.ndpbeta.nla.gov.au) was released to the public in August 2008 as a "beta" version. Users can now search and browse over 6.8 million articles from over 720,000 newspaper pages dating from 1803, with more pages regularly being added.

Through this service the NLA is providing access to every article, advertisement and illustration on every newspaper page being digitised. Users can browse the newspaper pages or search across the full text of the articles. Key users of the service range from academics, family historians and social and economic researchers through to school students.

As at the end of October 2009 the NLA has received a significant amount of feedback from a range of users of the service, including an Australian musicologist who has described the service as providing a "momentous advance" in his research focussing on a composer in Sydney in the 1840s. Another has said: "I am a historical researcher and this site makes me want to leap up and down with excitement because this is an amazing resource which will save me an enormous amount of time."

Angus Trumble, Australian scholar and author of a number of books and articles on an impressively wide range of topics, says in his blog *The Trumbel [sic] Diaries*: "I cannot think of any resource for the study of Australian history that has in my lifetime come as close to providing almost overnight such an enormous sweep of access."[1]

Trumble continues:

[I]f you want to monitor the imminence of the colonial reputation of Queen Maud of Norway, or find and document the career of a particular racehorse, or study the incoming shipments to the Sydney agents of certain Paris milliners…or find someone who has hitherto been completely invisible to most if

---

[*] Rose Holley managed digital projects at the University of Auckland and held a number of roles on the New Zealand National Digital Forum before moving to the National Library of Australia in 2007 to project manage the establishment of the Australian Newspapers Digitisation Program and service. She is a member of the RLG Working Group on Social Metadata and is passionate about providing wider access to resources via digitisation and giving users the technology to interact and engage with digital resources.

[1] Trumble A, "Australian Newspapers beta" (*The Trumble Diaries* blog, 22 May 2009): http://www.nla.gov.au/ndp/news_and_events/media_releases/documents/TheTumbrelDiaries_22May2009AngusTrumble.pdf viewed 4 November 2009.

---

not all other published sources – *Australian Newspapers beta* will henceforth need to be your first stop.

The NLA is providing a service that engages with communities that use newspapers as a key research tool by enabling interaction with, and enhancement of, the content. Features built into the service include the ability for users to add personal comments to articles, add tags and also correct the electronically-translated text created through the automated process of Optical Character Recognition (OCR). To date this is the only known free online newspaper service in the world that provides this level of user interactivity.

Throughout the development of the ANDP one of the strongest philosophies has been to ensure transparency of processes and sharing of lessons learned with the broader library and user community. To this end, the ANDP has been active in ensuring documentation, such as progress reports, specifications, workflow and process diagrams, is made available from the ANDP website (http://www.nla.gov.au/ndp/project_details).

Overall the ANDP is delivering an important information resource for all Australians. The Australian Newspapers service revolutionises access to newspapers and aims to provide a single point of access for all Australian newspapers, freely available via the internet.

THE VALUE OF HISTORIC AUSTRALIAN NEWSPAPERS

Early Australian newspapers are one of the most important resources that provide contemporary accounts of how the colonies were governed and of key historic events that shaped the nation. They reflect the day-to-day lives and circumstances of our ancestors and are a significant record of the social, political, economic and cultural issues of the time. This is portrayed not only in the written articles but also in the images, advertisements and even the headlines and layout of the newspaper. It is for these reasons that newspapers are heavily used in Australian research libraries to support historic enquiry.

Australia's first newspaper was the *Sydney Gazette and New South Wales Advertiser*, first published on Saturday 5 March 1803. It was a government gazette published by authority of the Governor of New South Wales with the important role of distributing official announcements, shipping news, excerpts from foreign newspapers, and local social news. It also contained invaluable information about the convict colonies, including convict movements, punishments, escapes and releases. In each of the other Australian colonies, the first publication was also a government gazette. These early papers are critical for genealogists trying to find details of convict relatives. Later papers give information on topics of interest such as the gold rush. By the end of the 19th century a number of metropolitan, provincial and suburban newspapers were being published and weeklies were starting to appear. These newspapers all played an important role in reporting news from abroad as well as the recording of Australian daily life.

Researchers from varying fields of enquiry have discovered the digitised newspapers in the service and are using it to assist with their research. Many have already found hitherto unknown information, such as the Light Railway Research Society of Australia (LRRSA). One member says:

> The service has generated a great deal of interest, discussion and enthusiasm among the 250 LRRSA members. For example, researchers John Browning and Phil Rickard have been looking for information on sugar cane mills and tramways in Queensland, and early tramways and tram roads in the Northern Territory and Tasmania. Using ANDP, they have found references to previously unknown tramways. They have also located newspaper articles about the early use of timber tramways in Queensland around the 1860s and information about industrial railways in Tasmania that were operating almost 40 years before the first steam public railway opened in the state in 1871.

Dr Bruce Moore, Director of the Australian National Dictionary Centre at the Australian National University, has been tracing the development of the Australian accent and of Australian vocabulary, linking this to major movements in Australian history and culture. He and his colleagues are now determining words that originated in Australia – so-called "Aussie lingo". Dr Moore says, "We can now search Australian Newspapers online and easily find occurrences of words we are trying to verify as being first used in Australia".

Professor Ian Fraser, immunologist, medical researcher and Australian of the Year in 2006, is using the service to research the history of influenza in Australia. A government-funded research

project, with which the NLA is collaborating, is using the service to study climate change and weather in Australia up to 1900. Examples of use by other researchers are detailed in a recent *Gateways* article.[2] Genealogists in particular have found the newspapers invaluable in their research. As one says, "It is the best thing that has ever happened to me in 20 years of family history research".

THE AUSTRALIAN NEWSPAPER PLAN (ANPLAN)

Prior to the ANDP, researchers wanting to use newspapers had a tedious and difficult task ahead of them. There is a not a complete central collection of newspapers in Australia and researchers would have to travel around the country to access microfilm or hard copy newspapers in each State or Territory library, with some newspapers held at the NLA. Only a few of the newspapers have indexes (eg *The Argus Index* (http://www.nla.gov.au/argus)) so searching for subjects, names or keywords was largely impossible. Instead many researchers would spend hours browsing randomly across microfilm reels.

Fortunately, however, since 1992 Australian libraries have been taking a national, collaborative approach to preservation and access of newspapers through the Australian Newspaper Plan (ANPlan, http://www.nla.gov.au/anplan). ANPlan members comprise the National, State and Territory libraries and their broad objective is to establish and work to a national plan, which coordinates microfilming and preservation activity at the national level. Through ANPlan each partner library has responsibility for collecting, preserving and providing access to each newspaper title published in their particular jurisdiction. This aims to ensure that at least one hardcopy of every newspaper published in Australia is retained in a library collection for as long as possible and that a surrogate copy, such as microfilm, is made available to facilitate long-term public access at the national level. ANPlan creates "master" microfilm copies of significant newspapers. Reading room copies can be made from master copies and these are usually purchased by any library that wants to give access to the newspaper title. Only the master copy can be used for digitisation purposes. The plan gave the ANDP a head start since it meant that most of Australia's significant newspapers were already microfilmed for preservation and therefore the master copies were available for the digitisation program. Digitising from microfilm rather than hard copy newspapers has significant cost savings.

The key objectives of ANPlan are as follows.

Collecting

In the area of collecting, partners are required to:
*   collect hardcopies of all newspapers from their area of responsibility as published; and
*   identify, locate and collect missing titles and issues.

Preservation

The second area of responsibility for ANPlan partners is preservation. Partners are required to:
*   retain as long as possible one hardcopy of every newspaper published in their jurisdiction;
*   create or purchase an archival standard master reproduction microfilm and at least one working copy microfilm reproduction of every title; and
*   provide appropriate housing and management of all copies of every title.

Access

The third area of responsibility for ANPlan partners relates to access. Partners are required to:
*   catalogue all print and microfilm holdings of newspapers into the Australian National Bibliographic Database (ANBD) through Libraries Australia; and
*   provide easy access pathways to the content of each title.

It is through the collaborative work of Australian libraries and under the auspices of ANPlan that the ANDP is also helping to achieve the overall objectives of collecting, preserving and providing access to Australian newspapers.

---

[2] Holley R, "Researchers have a Field Day with Digitised Historic Australian Newspapers" (2009) 101 *Gateways* 200, http://www.nla.gov.au/pub/gateways/issues/101/story03.html viewed 4 November 2009.

AUSTRALIAN NEWSPAPERS DIGITISATION PROGRAM (ANDP)

The ANDP began in November 2006 when the National Library of Australia received $10 million to digitise 4 million newspaper pages over the next four years. Ten years earlier the NLA had led a cooperative project to digitise newspapers: the Australian Cooperative Digitisation Program (ACDP, http://www.nla.gov.au/acdp). The ACDP gave several pointers for the way forward and one of these was that any future project would need to be a national collaborative led by the NLA. This was for two reasons; first, newspapers are distributed; and secondly, the NLA was the only library with the capability to develop a national storage and infrastructure system and delivery system. This would be a costly, difficult and large exercise, requiring expert staff and significant funds. Hence in the 10 years following ACDP, the project was scoped more than once and grants were applied for unsuccessfully until 2006.

The aim of the ANDP is to develop one national access point for all Australian digitised newspaper content, and a centralised storage system for the files. These objectives have been achieved. In the initial phase of the ANDP (2006-2011), one major daily newspaper title from each Australian State and Territory was selected for digitisation by State and Territory libraries. The earliest title is the *Sydney Gazette and New South Wales Advertiser*, the first newspaper to be published in Australia. The selected titles are being digitised from the date on which they were first published up until the end of 1954, or when the newspaper ceased publication. From 1955 onwards copyright applies and digitisation of newspapers published after this period may be undertaken in future if permission is obtained from the relevant newspaper publishers.

As the program has progressed so successfully over the past two years and the original estimate of number of pages in a title was over, rather than under, estimated, additional titles have been selected for inclusion, taking the total title count to 93. These additional titles will enable the NLA to achieve digitisation and delivery of around 4.4 million newspaper pages (40 million articles) by mid-2011. A full list of the newspaper titles currently selected for inclusion in the ANDP can be found on the public website.

As well as the selected newspaper titles being funded by the National Library of Australia, in December 2007 the Vincent Fairfax Family Foundation provided the NLA with $1 million in additional funding for inclusion of the *Sydney Morning Herald* in the ANDP. As Australia's longest running daily newspaper this nationally significant title is an important and welcome addition to the ANDP.

Through the ANDP, the NLA is also developing a model for national collaboration and contribution. At present the NLA is funding and managing the newspaper digitisation activities; however, it is keen to enable participation by the wider community. It is anticipated that in the second phase of the ANDP, contribution of local and regional newspaper titles by libraries and other institutions will be supported. For example, if a public library or local historical society has digitised newspaper titles themselves, it is desirable that this content is able to be integrated into the ANDP in the near future.

The public website for the ANDP includes a comprehensive range of information to keep the public and stakeholders informed – for instance, standards and workflows used; systems architecture, coverage and progress; development plans; public presentations; and outcomes of research.

WORKFLOW OVERVIEW

The entire workflow process from digitisation of the newspaper page through to delivery of the digital page to the user is complex. The following steps provide details of the current approach.

Identify and locate selected newspaper microfilm

In order to efficiently and cost-effectively make available a large volume of digitised newspaper content, the ANDP is creating digital page images from microfilm versions of the selected titles. As the NLA does not own the majority of the microfilm required for the digitisation process, the

microfilm is being sourced and borrowed from the Australian State and Territory libraries, as well as a Sydney microfilming bureau, W&F Pascoe Ltd, who owns the microfilm for a number of Australian newspapers.

Digitisation

The creation of the digital page images from microfilm is undertaken by external contractors based in Australia. As at the end of October 2009, nearly 2.6 million digital page images have been created.

Quality assure digital images

Once the digital images have been created by the external contractors they are delivered to the NLA where they are quality assured. This process involves checking that all page images are present, in the correct order, the right way up, cropped correctly and split correctly (where more than one page is on a single microfilm frame). Missing and duplicate pages or issues are identified and flagged.

Content analysis and Optical Character Recognition processing

The digital newspaper page images are then sent to other external contractors based in India where content analysis and OCR processing is undertaken. This is the most complex part of the entire workflow and involves:
- Zoning each page into areas (identifying each individual article and/or illustration on the page).
- Identifying and linking those articles together that continue across pages and linking any illustrations to the relevant article.
- Applying a category to each identified article (news, family notices, advertising and detailed lists, results and guides).
- Converting the newspaper page images into a full-text searchable file using OCR software (Abby FineReader).
- Rekeying of identified parts of the OCR text for each article. As the quality of the OCR, or "electronically-translated text", can vary greatly, the article title, subtitle, author and first four lines of the article text are rekeyed. This means that articles are more easily retrieved and the results of keyword searching are more accurate for users.

Quality assure processed digital images

Once the digital images have been through content analysis and OCR processing, the output files are returned to the NLA as XML files using ALTO and METS metadata exchange formats. Further quality assurance is undertaken to ensure that the external contractor has delivered files that meet the NLA's required specifications.

Public availability

All completed and quality-assured digital newspaper page images are made available through the Australian Newspapers service. Through the service users are able to browse the newspapers or keyword search across every article, advertisement and illustration on every newspaper page. The service is free and internationally available via the internet.

Data enhancement by the public

The public have the ability to help improve the quality of the OCR text by correcting text as they read it. A simple intuitive system for text correction is available while viewing newspaper pages and text, and this has been heavily used to date. Public users have improved 6 million lines of text as at the end of October 2009. Users may also add tags and comments to newspaper articles and these features have been heavily used as well. The data improvements are kept in separate layers, in addition to the library-generated content. The original content is not overwritten. All layers are searchable.

INFRASTRUCTURE AND SYSTEM ARCHITECTURE

Figure 1 shows a simplified diagram of the infrastructure. At the core of the system is a MYSQL database and a Lucene search index. The NLA has an online working space of 63 TB (soon to be increased), which is used for newspaper pages partway through the workflow process. Images partway

through the workflow need to be kept online in order that: quality assurance can take place; images can be moved from reel jobs (microfilm) to batch jobs (OCR); file names can be changed from sequential numbers to persistent identifiers; image manipulations can be applied to the image pairs; and so that derivatives can be created. Once an image has been completely processed it is moved to master storage offline. Master storage space is unlimited and LT02 tapes are used. The NLA currently has the storage, infrastructure and staff capacity to process 1 million newspaper pages per year, and this may be increased in the near future.

FIGURE 1   ANDP Infrastructure
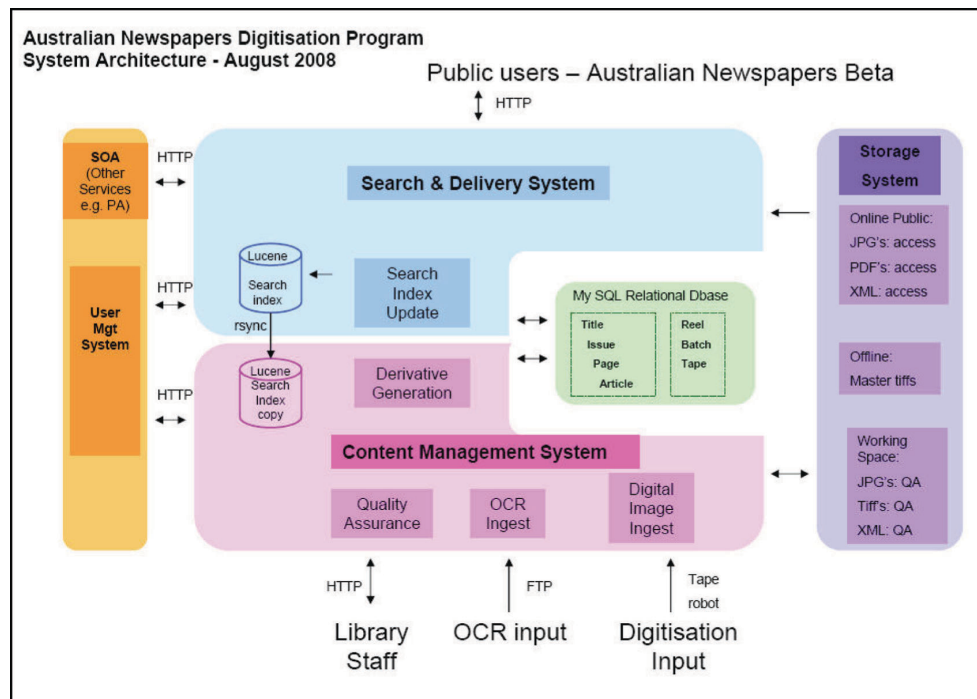
The open source code and system specifications for the Search and Delivery System are publicly available at the Library Forge website (http://www.code.nla.gov.au) and the Newspaper Content Management system (with Quality Assurance modules) will shortly be available as well.

AUSTRALIAN NEWSPAPERS PUBLIC DELIVERY SERVICE

The Australian Newspapers service is truly innovative and unique in the way in which it delivers digitised newspaper content and engages with the online user community. The ANDP has embraced Web 2.0 technology in order to provide a cutting edge service that allows users to interact, contribute and add value to the newspaper content. There is currently no other equivalent newspaper service in the world that allows users to tag, add comments and correct the electronically-translated text.

The search and delivery system was developed in-house by the ANDP IT team using open source software, as there were no suitable systems in the open marketplace that would allow the NLA to fully meet the objectives of the ANDP. The team first looked at sites that enabled viewing of large quantities of data using different techniques, and interactive sites using Web 2.0 before beginning to develop the NLA's own newspaper viewer and online service. The newspaper viewer uses technology from Google maps to zoom and navigate around the data. The team was keen to meet undefined user needs for using and interacting with the data. Therefore the Google approach to software development was followed and at a very early stage a prototype was released to stakeholders and users, shortly followed by a beta version. The software development followed the agile process rather than a waterfall model. Users were strongly encouraged to give feedback that would shape the development of the future

service and this they did with great enthusiasm. At a later stage, structured usability testing was also undertaken with the public to ensure that the system was intuitive and easy to use. Development was strongly focused on the needs of users and the ANDP IT team were encouraged to think innovatively. It was this approach that resulted in the idea of public text correction being first discussed and then implemented. The beta service was available to the public from August 2008 to August 2009. During the first six months thousands of users found the service and suggestions for enhancements were received from 600 users. The response was overwhelmingly positive with many stating that the service already exceeded their expectations. Development and feedback continued and in August 2009 the beta status was removed. Significant views held by the majority of users were that:

- imperfect data still holds huge value – a part of a newspaper page, or incorrect OCR is better than getting nothing at all;

- quantity is more important than quality – having the entire run of a paper even if it is in poor condition is better than having only the best pages available; and

- being trusted to correct the text and being kept informed about plans for the service builds loyalty and commitment in the user community.

The NLA believes that development of the beta service has not resulted in a traditional library database, but rather is providing users with innovative ways of exploring full-text resources. The interface includes features such as relevance ranking and clustering of result sets, eg grouping of results by newspaper title, article category, date range and article word count. This is important when keyword searches may return thousands of results. Perhaps the most innovative feature is that which allows users to correct the OCR or "electronically-translated text". While OCR works well for documents with a consistent typeface and standard format, the nature of historical newspapers with varying fonts and print quality, as well as high article density with little white space between text, means that OCR accuracy is often low.

The ANDP team undertook research into how OCR accuracy could be improved technically before it came to the conclusion that manual intervention was still the most effective means.[3] As the human eye is much better at reading text correctly than the OCR software, the Australian Newspapers service allows and encourages users to correct errors in the electronically-translated text. The contributions that users make in correcting the text add value to the service and improve searching for subsequent users. It should be noted, however, that history cannot be changed with one deft keystroke, as the original OCR text is still retained and remains searchable in the database, in addition to the digital newspaper page image remaining a true surrogate of the original.

Other interactive features of the service include the ability for users to add comments or annotations to articles. These can contain additional information about the subject of the article, or can inform other users that the information contained in the article is incorrect, eg "The writer would appear to be Corporal John Heaney No 91, 6th Bn, who was wounded in the left leg at Gallipoli on 14 July 1915" and "Convict ship 'Captain Cook' incorrectly shown in paper as the 'Captain Cooke'". These comments can be made available for all to see or, if a user is registered, can be added as private study or research notes. Users can also tag articles with relevant keywords relating to the subject or content of the article.

In the homepage of the Australian Newspapers service (Fig 2) the user has the ability to search articles using keywords, phrases and dates or to search for a particular issue by selecting the newspaper title and date (year, month and day). At the foot of the homepage, users have the option to sign in, the five top text correctors are identified, and recent comments and tags are shown.

---

[3] Holley R, "How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs" (2009) 15(3/4) *D-Lib Magazine*, http://www.dlib.org/dlib/march09/holley/03holley.html viewed 4 November 2009.

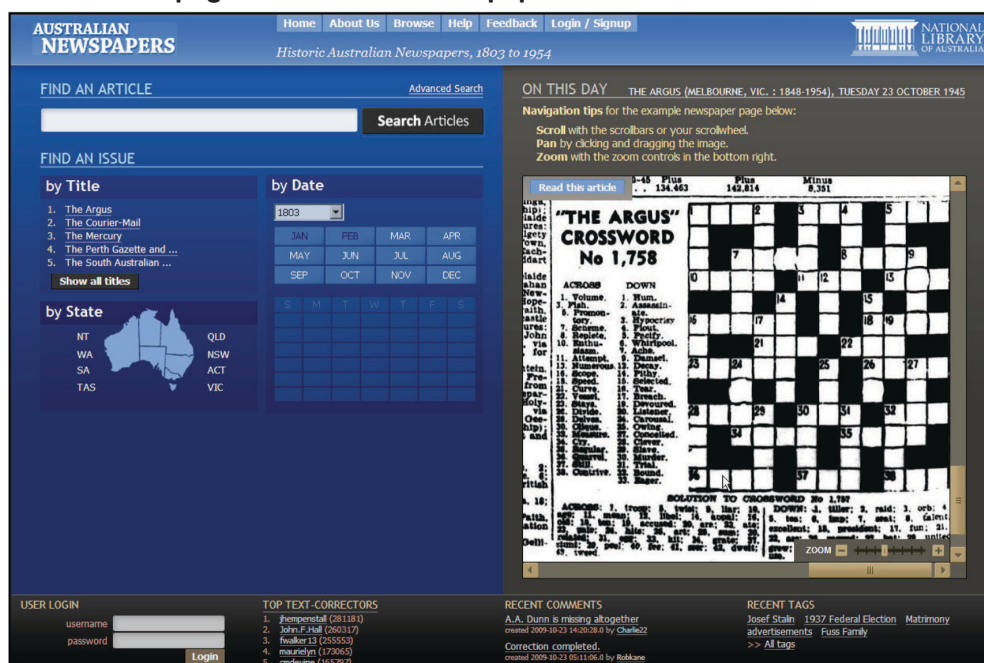**FIGURE 2   Homepage of Australian Newspapers service**

Figure 3 shows a results page from a keyword search for Pharlap (Phar Lap), a famous Australian racehorse. The results returned are sorted by relevance and can be sorted by date. Results can be refined (as shown on the left) by newspaper title, or by category of article, whether or not it is illustrated, by decade and by word count. Where a tag has been assigned to an article it appears as a link on the article and the tag can be searched by activating the link. If the article has been corrected the date of last correction is also shown.
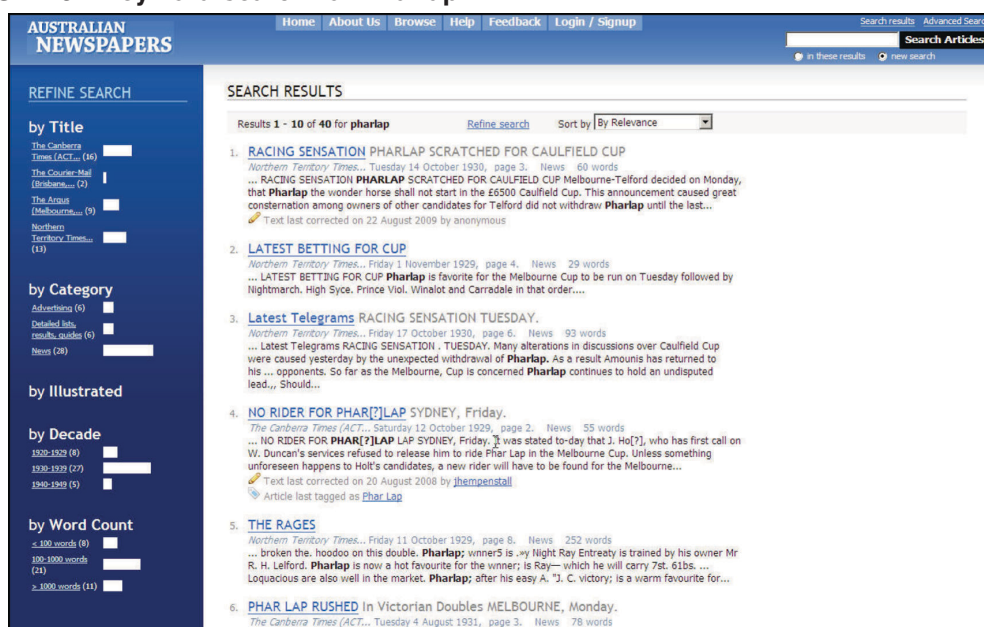
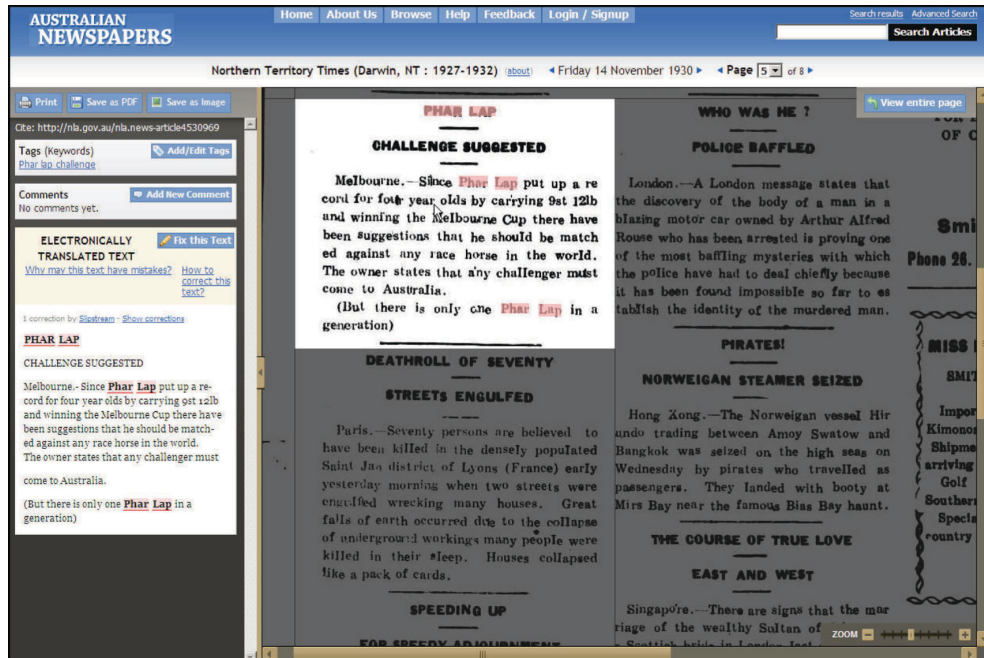**FIGURE 3   Keyword search for Pharlap**

Figure 4 presents the article in the context of the page, with search terms highlighted and with the user interaction functions – tagging, text correcting and comments – presented on the left side. Each article is assigned a Permanent Identifier (PI) for citation purposes.

FIGURE 4 Article view

PUBLIC TEXT CORRECTION

Since release of the Australian Newspapers service just over a year ago, over 6 million lines of OCR text have been corrected by about 5,000 users. One exceptional user has corrected over 280,000 lines of text in over 7,000 articles. This has surprised and pleased not only the NLA but also the international library community. The experience is now shaping future thinking for social engagement and data enhancement activities at many libraries around the world, particularly in relation to full-text digital materials.

A risk analysis was carried out before text correction was implemented. Technical issues such as moderation and roll back were also discussed. No moderation module was implemented.

FIGURE 5   Text correction view
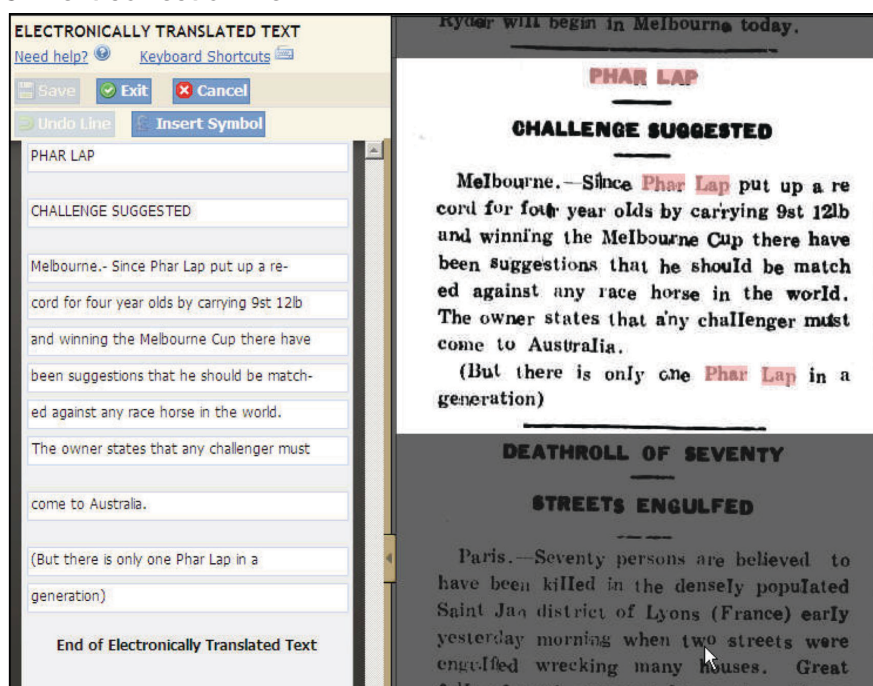
Research on text correction activity in the Australian Newspapers service, a survey of the top text correctors and resulting report were published in March 2009, titled *Many Hands Make Light Work*.[4] The following sections give a brief overview of some of the key points from the report and update the statistics.

Identified benefits of public text correction
- Improved data quality and keyword searching for all users.
- Meeting user expectations of a quality service.
- Building new virtual user communities and social networks based on the collection.
- Gaining experience in innovative use of Web 2.0 technologies and assessing their potential for other discovery services.

Identified risks of implementing public text correction
- No models to follow; unknown how technically difficult it may be.
- Potential vandalism of text.
- Large amounts of text correction activity could compromise service performance.
- Not sure if users may take up text correction so development time may be wasted.

User activity
- Many users found the text correction rewarding and also addictive.
- Users were actively correcting much more text than had been expected.
- Top correctors spent up to 40 hours per week on text correction.
- Eighty per cent of the correction was done by users who had chosen to register and the other 20% by anonymous users.
- The super correctors (top 10) did 29% of the text correction with other users (approximately 5,000) doing the rest.

---

[4] Holley R, *Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers* (National Library of Australia, 2009), http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf viewed 4 November 2009.

- The top 10 correctors were the same for the first year.
- Most correctors are motivated by the thought of doing something worthwhile that helps others, and/or being able to help in accurately recording Australian history. Addition of new content and acknowledgement in the "hall of fame" (top correctors each month) is an added motivator for some text correctors.
- Half of the text correctors are correcting alongside their research and the other half simply view it as an interesting voluntary activity.
- Some text correctors work methodically through entire articles, while others only correct the odd word here and there – usually personal names and places.
- Within the first six months an active text-correcting community of 1,300 people developed who had corrected 2 million lines of text. This rose to 5,000 people and 6 million lines of text at the end of 14 months.
- Some correctors want to be given a list of topics or articles to correct to provide them with goals.
- Some correctors want social networking features so they can communicate and interact with each other.
- Some correctors want to take a more active and responsible role as moderators or coordinators of digital volunteers.

Lessons learned

- Text correction is better defined as data enhancement since no data is overwritten but rather layers are added.
- If the public are given a high level of trust and responsibility they will respond with loyalty and commitment.
- The public want to help libraries make resources more accessible and add value to them. They want to interact with data and each other and can do this easily with Web 2.0 technologies if applied across library collections.
- There is huge potential for applying text correction and also text transcription technologies across library and archive collections in the future.

FIGURE 6   Profile of super correctors (top five)

	Julie	Lyn and Maurice	Mick	Catherine	Fay
Interests	Family/local history	Family/local history; shipping	Family history/early Australian history	Doing something to help other people	Family history research
Age and status	31-45; stay at home mum	55-62; retired couple	41-60; retired	31-45; working full-time	61-80; retired
Activity	15-45 hours per week	15 hours per week	12 hours per week	15 hours per week	Varies
Why do it?	Enjoys it; a great way to learn about history; a service to the community	Sick of doing housework!	It benefits me and other people	Want to do something useful; finds the content fascinating	Need something to do in my spare time; it benefits me and others and I enjoy the challenge
Will you continue?	Yes – a "must do" mission	Yes – it helps us and other people	Not sure	Yes	Yes
What would keep you motivated?	More papers added	Working on specific projects/topics	More papers added	Being given ideas on topics to correct	?

FIGURE 7   Comments from public users about text correction

"I like being able to see the OCR text and the original side by side. This is a smart move to entice people to correct text."
"OCR text correction is great! I think I just found my new hobby!"
"An interesting way of using interested readers' 'labour'! I really like it."
"A wonderful tool – the amount of user control is very surprising but refreshing."
"A great idea. I would sooner have online access, and search possibilities, even with the mistakes, and am happy to correct the scanned text as a quid pro quo."
"Would like to say this is a great initiative although I think there should be a warning about using this site and its possible addictive effects! I have a great deal of trouble getting back to what I should be doing at times."
"Wow – well I got sucked in! I can see why everyone is editing the OCR text…it's compulsive!"
"The decision to let members of the public edit & correct articles must have been a difficult one, but it's a great idea."
"Thank you! You lot are so cool!"
"While going through a whole month in a slightly obsessive crazed mind searching Australian newspapers beta online, I just realised the kilos I've stacked on in just one month. I can't seem to snap out of it; from dawn to dusk I seem to be in this website craving to find more on my ancestors – all the gritty stories. Housework seems to have taken a backburner and meals are starting to come out of cans…is there an AA for genealogy junkies?"

STATISTICS SUMMARY

FIGURE 8   Usage of the Australian Newspapers service over the last 14 months and uptake of data enhancement activities

Statistic type	4 Aug 2008 (public release)	4 Nov 2008 (3 months after release)	4 Feb 2009 (6 months after release)	4 May 2009 (9 months after release)	20 Aug 2009 (1 year after release)	1 Oct 2009
Number of pages in service	73,000	367,000	367,000	367,000	538,334	721,290
Number of articles in service	730,000	3.5 million	3.5 million	3.5 million	5.8 million	6.8 million
Unique visitors to site	-	94,000	205,000	347,000	492,000	665,000
Number of registered users	-	1,488	2,994	3,796	4,762	5,498
Lines of text corrected	-	1 million	2.2 million	3.4 million	4.7 million	6.1 million
Number of articles corrected	-	60,000	104,000	154,000	216,093	276,339
Number of comments added	-	800	1,806	2,582	3,441	4,247
Number of tags added	-	18,000	43,000	73,733	105,028	166,572

FUTURE PLANS

The NLA has worked with Google to provide wider access to Australian historical newspapers via Google's News Archive (http://www.news.google.com.au/archivesearch). By implementing a harvesting site map that enables Google to index all of the articles, access to the digitised newspapers is made more widely available. When users search from the Google News Archive site, results from the Australian Newspapers service are also returned, with users then referred to the service to get the full article. The NLA has populated relevant Wikipedia articles on individual Australian newspapers with external references to the digitised content at title level. It has also dynamically pulled information from Wikipedia articles on newspapers into the Australian Newspapers service "About" pages at title level. This means that another access point has been created from Wikipedia. The NLA is interested in setting up harvesting sites or linkages with other significant access points in the future so that newspaper articles are discoverable from those places.

The NLA is currently working on a project to integrate the service into a single discovery service for Australian content. This new service – "Trove" (http://www.sbdsproto.nla.gov.au) – is still in the prototyping stage and is due for launch in November 2009. This service would give users the option to search across newspaper content only, or to do an integrated search across other content such as pictures, music, books, maps or research. This would be particularly useful for researchers who are interested in their subject, eg a person rather than the format of the resource. Initial tests during the beta phase, where users also had the option to view related resources such as pictures and published works retrieved from other NLA discovery services including Picture Australia (http://www.pictureaustralia.org) and Libraries Australia (http://www.librariesaustralia.nla.gov.au), received a very positive response. The user interactions piloted in newspapers, such as tagging and commenting, will also be enabled more widely in the rest of the NLA's collections because they have proved so successful.

During the next phase of the ANDP, the NLA will also implement a national framework to enable contribution of additional digitised newspaper content by other libraries and institutions. Another grant application is underway for $6 million to enable digitisation of regional newspapers, and the outcome of this application will be known shortly.

The NLA is now moving forward with mass digitisation of out-of-copyright Australian journals and books using the knowledge, software and workflows developed for newspapers. *The Australian Women's Weekly*, with the permission of the publishers, will be the first journal to be digitised (that sub-project commenced in August 2009).

A task not yet undertaken is the formal launch, publicity and promotion of the Australian Newspapers service. This will happen in the immediate future now that the service is no longer in beta version and significant development has taken place. A raft of enhancements as requested by the public will be implemented in early 2010. Most of these are related to enabling better user interaction and communication around tags, comments, user profiles and forums.

SUMMARY

A tremendous amount of work has taken place within the National Library of Australia over the past two years to progress the ANDP. This has largely been due to a small, high-achieving team of five people who took an innovative and energetic approach to what has been the NLA's biggest financial and most significant access project for some years. Without the full-hearted support of senior management, the cooperation and assistance of the State and Territory libraries, and a significant budget this project would not have been successful. As it is, the Australian Newspapers service is now the success story of the NLA and a welcome advancement to millions of Australians. The second part of the success is the willingness of the public to interact and engage with the service and make it even better.