

# MEMORANDUM



DATE 20 August 2008  
REF NLA07/1529  
TO ANPLAN  
FROM Rose Holley – Manager Newspaper Digitisation Program  
SUBJECT **Progress Report on the Australian Newspapers Digitisation Program March – August 2008**

---

## **Purpose**

This report summarises progress on the Australian Newspapers Digitisation Program (ANDP) over the last six months from March – August 2008 and outlines work to be completed in the next 6 months.

---

## **1. Work progress March 2008 – August 2008**

- Re-processing of the 50,000 page pilot data according to the Library's specifications was undertaken by the OCR contractor from April – June 2008. Production of further pages began in June and is ongoing. To date a further 50,000 pages have been accepted by the Library.
- The 'selected titles' list and the 'scanning schedule' are now available via the website [http://www.nla.gov.au/ndp/selected\\_newspapers](http://www.nla.gov.au/ndp/selected_newspapers). They are being updated every 6 weeks to show progress.
- The RFT documentation for 2 tenders (Scanning and also OCR services) was developed and then released on 25 July 2008. It is the Library's intention to engage a panel of contractors for each of these services.
- Further software development on the Search and Delivery system was undertaken based on the feedback gained from stakeholders about the Search and Delivery Prototype (released in December 2007).
- The Australian Newspapers Beta search service was released to the public and stakeholders on 25 July 2008. On release it contained 70,000 pages (approximately 700,000 articles). As pages are processed by the OCR contractor they will be added to the service. The Beta service will remain in place as software development continues and users are being encouraged to give the Library feedback over the next few months. There has been a significant positive response to the Beta service and several items have appeared in the media about it.

- Work continued on enhancing the Library's technical infrastructure to support the Newspapers Digitisation Program. Two new servers were purchased and installed. One to support the newspapers public search and delivery service and one to support the newspapers Content Management System.
- Software development continued on the Content Management System. A new module to enable quality assurance of work undertaken by OCR Contractors was developed.
- Scanning of microfilm and quality assurance (metadata added, pages sequenced, missing pages and issues identified, and batched ready for OCR) of data continued and to date 1.4 million pages have been digitised (scanned) and quality assured. The scanning rate has increased from 18,000 per week to 35,000 per week.
- A project plan to digitise the Sydney Morning Herald 1831- 1954 in tandem with ANDP activities was developed.
- The existing microfilm of the Sydney Morning Herald was digitised and quality assured (approximately 500,000 images).
- Project documentation on the public website [http://www.nla.gov.au/ndp/project\\_details](http://www.nla.gov.au/ndp/project_details) was updated and new items added.
- ANDP staff (and some stakeholders) attended the IFLA International Newspaper Conference in April 2008 in Singapore. This provided an excellent opportunity to meet staff involved with digitising newspapers from other national libraries and Vendors/Contractors of related services.
- Presentations on the ANDP were given at the National Folk Festival and to National Library Petherick readers. One to one sessions for visitors from other national libraries (attending the IIPC Conference), the Australian Museum, The Canberra Times, The Vincent Fairfax Family Foundation, and Fairfax Media were also given.

## **2. Work to be completed September 2008 – February 2009**

- Evaluation of 2 RFT's (OCR and Related Services and Scanning Services) and appointment of panels of contractors for each.
- Work will continue on digitising (scanning) newspaper pages. 3 million pages are yet to be scanned.
- Work will continue to OCR process newspaper pages. 1.3 million pages have been scanned and are awaiting processing.
- In consultation with stakeholders develop a model for national contribution of additional newspaper content. This will involve development and distribution of guidelines for contributors, and development of workflows and infrastructure at the National Library.
- Modification of the Content Management System to enable Contribution of content by stakeholders.

- Feedback from users of the Beta service will be collated. Reported bugs will be fixed. Further development of beta into Search and Delivery 'version 1' will occur in 2009.