

MEMORANDUM

DATE 14 February 2008

REF NLA07/1529

TO ANPLAN

FROM Rose Holley – Manager Newspaper Digitisation Program

SUBJECT **Progress Report on the Australian Newspapers Digitisation Program September 2007 – February 2008**

Purpose

This report summarises progress on the Australian Newspapers Digitisation Program (ANDP) over the last six months from September 2007 – February 2008 and outlines work to be completed in the next 6 months.

1. Work progress September 2007 – February 2008

- The National Library has completed development of the software required to support the workflow process, and the content management system.
- A search and delivery prototype has been developed by the National Library. This was released to stakeholders for testing and feedback in November 2007.
- Specifications for derivative images for use in the search and delivery system have been developed and the required derivative images (JPEG, PDF) have been generated from the master TIFF files.
- 700,000 pages have been converted from microfilm into digital images and these have been quality assured by the Library (metadata added, pages sequenced, missing pages and issues identified, and batched ready for OCR).
- Digitisation is ongoing with microfilm being scanned at a rate of 18,000 images per week.
- The OCR contractor completed processing of the 50,000 page pilot data however the output files did not meet the Library's specifications. The contractor has revised specifications and is reprocessing the pilot data. This has resulted in a delay in the Library's timeline for making the service publicly available.

- In September 2007 the Library received \$1 million from the Vincent Fairfax Family Foundation to support the digitisation of the Sydney Morning Herald from its inception in 1831 to 1954. In addition, the National Library will contribute \$700,000 towards this Project. The inclusion of the Sydney Morning Herald in ANDP had previously been deferred due to the quality of the existing microfilm i.e. missing and damaged pages.
- Documents outlining current workflows and digitisation information were created and made available to ANPlan members and the public on the ANDP website <http://www.nla.gov.au/ndp/>
- Presentations about the program were given at:
 - Spydus User Group meeting (31 October 2007)
 - Australian Academy of the Humanities Symposium (16 November 2007)
 - Australian Media Traditions Conference (22 November 2007)
 - VALA (4 February 2008)
- Visits to the National Library to view the workflow processes and in-house developed software for the ANDP were undertaken by State Library of Victoria (Shirley Firth and Adrian Flint) and State Library of Northern Territory (Diana Richards).

2. Work to be completed February 2008 – August 2008

- Develop a project plan to digitise the Sydney Morning Herald 1831- 1954 in tandem with ANDP activities.
- Finalise the selection criteria for the ANDP and make the criteria available to stakeholders and the public via the website.
- Provide the selected titles list and the scanning schedule to stakeholders and members of the public via the website.
- Re-processing of the 50,000 page pilot data to the Library's specifications by the OCR contractor. When pilot data is accepted the Library can move into production phase (OCR processing of 500,000 pages).
- 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.
- Release RFT's for content conversion services (digitisation) and content analysis and Optical Character Recognition (OCR) services. It is the Library's intention to engage a panel of contractor for these services.
- Further software development based on the search and delivery prototype feedback from stakeholders and then public release of the Search and Delivery System 'version 1' for user testing. It is anticipated that the service will be publicly launched in late 2008.