THE AUSTRALIAN NEWSPAPERS DIGITISATION PROGRAM (NDP)

Rose Holley – Manager Newspaper Digitisation Program
Presentation at the Association of Parliamentary Libraries of Australasia Conference
26th July 2007, Australian Parliament, Canberra
Current Status

- 29 November 2006 approval from Minister for Arts and Sports for National Newspaper Digitisation Program

- Budget approved -$8 million for 3 million pages over 4 years

- 1\textsuperscript{st} March 2007 – contract signed with OCR supplier

- 2\textsuperscript{nd} April 2007 Manager for NDP starts and program enters initial phase
Content and Coverage

National Content

Initially a title from each state

Focus on major titles from each state first

Anticipated that ‘regional’ titles may be contributed later

Coverage: published between 1803 – 1954 (out of copyright)
Process in brief

National sourcing of selected newspaper microfilm masters.

Masters scanned by W & F Pascoe, Sydney to tiff files.

NLA perform quality assurance, add metadata.

Apex Publishing, India process tiff files - OCR, zoning, xml markup.

NLA QA files, ingest to system, create derivatives for delivery.
Timeframe

April - September 2007

✓ Pilot data to Pascoe’s and to APEX to design and test workflows, systems and software against agreed project spec.
✓ Developing software and infrastructure in-house to support workflows.
Timeframe

October 2007 – 2008

- Development of search and delivery software
- Production phase 1 begins – 500,000 pages in first year
- Public launch of service
- Progressive addition of content
Technology - internal

Old newspapers being processed and delivered using latest digital technology

• NLA developing in house:
  – Ingest and storage system
  – Workflow and content management system including quality assurance module
  – Search and delivery package

• NLA providing:
  – System Infrastructure
    (storage, backup, disaster recovery)
Technology - external

- Scanning microfilm using newest scanner (Flexscan) and software (nextstar) from NextScan [www.nextscan.com](http://www.nextscan.com)
Technology - external

Apex ‘Izaac’ software:
- Zoning areas and articles on a page
- Flag continuing articles across multiple pages
- Categorise articles on a page
- OCR text on a page
- Re-key headings and first 4 lines of text.
- Deliver xml files (alto) and METS files.
The Maitland Mercury and Hunter River General Advertiser

Vol. II. Saturday, January 20, 1844.

Job and Book Printing.

A new extensive addition having been just made to the Plain and Fancy Jobbing Office. The specie is now ready for business, and the work of printing may be carried on without any delay.

Maitland Mercury General Printing Office. Parties requesting Posts or Hand Bills, Books, Pamphlets, Catalogues, Letters, Title Deeds, Circumstantial, Gauchy, Bills of Lading, Cheques, Bills or colored Labels, Baker's Bills, Impressing, Law and other Forms, etc., can have them translated in a superior manner, with great dispatch, and at reasonable prices.

Jews for Marietta, Maribyrnong. Two Jews for Marietta, Maribyrnong, under engagement with the Sydney Jewish Company, will be ready to receive goods on or about the 21st January, 1844, and will meet with immediate dispatch, so as to have all their orders complete and seasonable.

To Printers.

A short notice to printers of the domestic and foreign press, that the present is the time to consider the necessity of having their business conducted in a regular manner, and that they should take measures to ensure the promptness and accuracy of their work.

To the Directors of the Maitland and Hunter River Agricultural Society.

A notice to the Directors of the Maitland and Hunter River Agricultural Society, to meet at the Maitland Mercury Office on Thursday, the 20th January, 1844, at 11 o'clock in the morning, for the purpose of considering the state of the society and the measures to be taken for its improvement.

Maitland Price Current.

A list of prices for wool, cotton, and textiles, as well as the current market prices in the Maitland district.

HUNTER RIVER AGRICULTURAL SOCIETY.

A notice to the members of the Hunter River Agricultural Society, that a meeting will be held at the Mercury Office on Thursday, the 20th January, 1844, at 11 o'clock, for the purpose of considering the state of the society and the measures to be taken for its improvement.

Several new members were then proposed and seconded, and the meeting adjourned until the following week.
Pilot categories (8)

News
News and current affairs including law courts and crime, official appointments and notices (e.g., Gazettes), commerce and business news, sporting news, social news

Advertising
Classified advertisements and notices, display advertising

Birth Death Marriage notices
Births, deaths, marriages, anniversaries, etc. notices

Obituaries
Obituaries
Pilot categories

**Editorial commentary and letters**
Editorials, leaders, letters and correspondence (usually to the editor - unpaid), editorial or political cartoons

**Shipping News**
Shipping news or intelligence

**Arts and leisure**
Art, literature, music, theatre, comics, shows, gardening, travel

**Detailed lists, results, guides**
Detailed sporting results, guides, radio and television guides, weather forecasts and results, election results, education results and courses, stock market lists, crossword puzzles.
Illustrations

Categorised and have a metadata flag to mark the illustrations

- Photo
- Cartoon
- Map
- Graph
- Illustration
Work in Progress this week…

- Derivative size and zoom technology testing
- Wireframes for search and delivery design
- In-house Quality assurance on images

Screenshots follow…….
Testing derivative sizes and zooming

Lorem ipsum dolor sit amet


berry, Strawberry, Pineapple, Lime Juice 1/2 bot.
Schweppes Lime Juice 2/4 bot.
Cordial 1/9 bot.
Raspberry Syrup 1/9 bot.
Lemon Cordial 1/9 bot.
Colonial Ox Tongue, 1 1/2 lb 3/9
Fiko Ox Tongue, 2 lb 5/3
Colonial Sheep Tongues, small 1/4
CANDIED PEEL—
Lemon 1/3 lb.
Orange 1/3 lb.
CITRUS 2/3 lb.
CAKES AND PUDDINGS—
Arnott’s Xmas Cake, 3 lb. 3/10 ea.
Ireland’s Xmas Cakes, uniced 4/6 ea.
Ireland’s Xmas Cakes, iced on top 5/3 ea.
Currants, 3 crown, extra choice 9d. lb.
Foster Clarke’s Custard Powder 6d.
12 oz. pkt. 1/2
1 lb. tin 1/9
Pickled Onions, Oakleaf, 1 1/6 bot.
Resella Chutney 1 1/3 bot.
Roast Spring Chicken, whole bird in jelly 3/
Resella Fruit Salad, in bots 1/6 bot.
SUNKIST and Del Monte, 2 lb. tins 2/6
Arnott’s best Fancy Biscuits, 2 lb. tins 2/3 and 3/9
Best Jordan Almonds 3d. oz. or 4/- 1 lb.
2 oz. tins Ground Almonds, 1/5 tin.
BURNETT’S EXTRACT—
Lemon, Raspberry, Kola, Lime Juice, Pineapple 1/- bot.
SWALLOW & ARIEL’S PLUM PUDDINGS—
1 lb. tins 1/-. ea.
1 lb. tins 3/- ea.
TILLOCK & CO’S PLUM PUDDINGS—
1 lb. tins 4/11-. ea.

Customers requiring Cakes and Puddings are advised to
Place their Orders early.
Mail Orders are packed free—But are freight extra.

NOCK & KIRBY LTD.
“FOR YOUR NEEDS.”
102 & 104, GEORGE STREET, SYDNEY.
Search and delivery wireframe


Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec hendrerit fringilla nunc. Pellentesque habitant morbi
QA system
In-house

Use 2 widescreen monitors placed vertically. Can view complete page within context of issue. Add metadata, sort out missing and duplicate pages within an issue. Prepare batches to send for OCR.
Contributions

• NLA working with State Libraries as part of ANPLAN.
• Feedback from State Libraries and stakeholders on prototype search and delivery interface.
• Tell your users about the program and which titles from your state have been selected for initial phase.
• Contribution of local newspapers at later stage.
The Australian Newspaper Plan

The Australian Newspaper Plan is an ambitious, ongoing project designed to collect and preserve every newspaper ever published in Australia, guaranteeing public access to these important historical records. There are similar plans for New Zealand’s newspapers.

Collect
Do you know the whereabouts of a missing newspaper?

Preserve
The overall aim of the ANPlan project is to preserve all Australian newspapers and to give the public access to them.

Access
ANPlan lets you access our newspaper heritage.

Partners
Keeping Up to Date

• E-mail contact - NDP@nla.gov.au
• Website: http://www.nla.gov.au/ndp/
Questions