

Scanning Hardware and Software for microfilmed newspapers

Author: Rose Holley (Manager – ANDP)

Version: 1.0

Date: 8 October 2007

The ANDP is only scanning from microfilm at this stage of the program, not from hard copy. If scanning from hard copy an overhead scanner would be required due to the fragility and large size of historic newspapers.

There are only a few models and makes of microfilm scanners on the market. In Australia there are a limited number of microfilming/digitisation bureaus who are able to undertake digitisation from microfilm. There is not therefore a great deal of choice about which scanner and software to use. Microfilm scanners cost in the region of \$100,000 and require expert operators so most organizations prefer to outsource digitisation, rather than buy equipment for use in-house. The National Library is outsourcing all the digitisation work for newspapers.

When considering what make and model of scanner to use it is important to find out what software and what version is being used on the scanner. The scanning software has a big impact on the outcomes of digitisation which must not be under-estimated. Key functions of the software are:

- Auto-recognition of microfilm frames
- Splitting of double/quadruple images in a single frame into single page images.
- Image manipulation to improve text for OCR
- Quality assurance/auditing of the microfilm reel

If any of these functions are not met adequately the result will be to receive images that are incorrectly or not split, having missing frames/pages from reels which may not be obvious, and the digital images when processed for OCR not having good results.

The contractor for the Australian Newspapers Digitisation Program is using NextScan hardware and software (www.nextscan.com). There are 2 microfilm models in use:

- Flexscan (scans microfilm and microfiche)
- Eclipse (scans microfilm only)

Both models have been upgraded to use the latest NextStar software. NextScan models are the preferred scanning hardware for the program because they are:

- durable, robust, easy to use and clean
- have a good scanning speed (scans an average reel in about 12 mins)

NextStar is the preferred scanning software for the program because:

- The software is continually being developed and enhanced (most other scanner software has had no serious development for the last 3 years)
- There is no 'frame-jumping' issue as with other scanning software
- The QA/auditing program in the software that allows automatic and manual checking of reels is very good

- The image enhancement algorithm in the software results in files that are very good for OCR, even when the original newspaper pages are very poor.
- The software is easy to use.

Full product information is available from the NextScan website www.nextscan.com There are 2 re-sellers based in Australia, and there are 4 scanning/digitisation bureaus in Australia that use the equipment (as far as the National Library of Australia are aware).

Product Highlight

NextStar



Introducing enhanced throughput and processing for 2007 – significantly faster and with more flexibility than NextStar in 2006 and other processing packages available in the marketplace today.

- Scan and display instantly the entire roll of film/whole fiche as one (ribbon) image (grayscale), regardless of image location, placement or density (vert/horiz, etc)
- Detect individual frames within the image for fast, accurate QA
- Verification process to ensure there are no missed images
- Detects and processes varying image size and density in one pass, on the fly
- No rescans or reloads needed
- Fastest throughput available in the marketplace - Won't slow down for even the largest images
- Simplified, intuitive interface - Minimal training required, so one operator can handle multiple scanners
- Audit function can be done separately, even remotely - this independent function optimizes workflow/reduced costs by eliminating potential bottlenecks.



NB: If local historical societies or other organizations are digitising from microfilm (whatever equipment is being used) the most important things are:

- To scan at the highest resolution available (400 dpi)
- Scan as greyscale (and save as bi-tonal as well)
- Use unique file name e.g. running number
- Split frames if necessary so one newspaper page is one image file
- To save the file as a tiff for preservation, re-use, participation in the ANDP (store on CD/DVD, portable hard drive if not enough server space).
- Please refer to the microfilm scanning specification on the ANDP website for more detail.