Now they know my name:
Creating a digital presence for Barton

Presentation by Andrew Stawowczyk Long
National Library of Australia

Australia’s cultural institutions hold the documentary and cultural heritage of the nation in trust for the Australian people, and it is one of their primary responsibilities to make their collections easily available. The Internet has become an ideal tool in delivering access. Early in 2000 the National Library of Australia decided to select a manuscript collection for digitisation, with the intention of making it available on the web. As 2001 marked the centenary of Federation, the decision was made to select the collection of papers of our first Prime Minister, Sir Edmund Barton. The Barton collection had also been identified as one for which there would be sufficient demand on-line. The generous donation of $50,000 from Mr Dick Smith and Mrs Pip Smith, when combined with the Library’s project budget, allowed work to commence in March. In addition to the primary goal of putting the collection on the Internet, the Library also wished to create digital surrogates for the Barton Papers. The Barton digitization project was the first project of such complexity and size ever attempted by the Library.

The Barton Papers were donated to the National Library by the Barton family between 1929 and 1996. They comprise of 16 security binders containing letters, 5 manuscript boxes and two folio boxes. Other items include a Royal Doulton plate, a gold medallion and a ceramic vase. All items in the collection have been divided into thirteen “series”. Initially, it was estimated that there were around 3,500 items to be digitised, but this proved to be wrong. Due to inaccurate finding aid and numerous items being more complex than we had realised (e.g. books with many pages), the final number of scanned images rose to over 5,500.

The majority of items in the collection are small. Only some 100 items are bigger than A3 page size (42cm x 29.7cm). Larger items, like newspaper cuttings, were scanned in segments and stitched. Oversized items that proved to be impossible to stitch, or three-dimensional items, were captured using high resolution, high quality Sinar Macroscan 23 and PhaseOne PowerFX digital cameras. The great variety of materials comprising the Papers sometimes presented a challenge to the scanning technician. Some pencil annotations on red-brown, blue or blue-green coloured papers were very difficult to read. Some of the correspondence material exhibits ink bleeding on the reverse side of the paper making it difficult to scan. Some items needed unfolding and to prevent damaging such documents a special platform was built by the Library’s Preservation staff around the scanner to keep them flat.

When the Barton Papers were transferred to the present National Library building in 1968, a finding aid or guide to the papers was created. Items were individually numbered and each piece was then indexed by author and recipient. Within the more important series, most of the items were described individually. For the purpose of digitisation,
work has been undertaken to expand series descriptions to item level for the whole collection, in order to accommodate persistent identifiers. This proved to be very tedious and time consuming undertaking. Paula Waring from the Library’s Manuscript Section found that ‘This process revealed many curious features of the arrangement of the collection that had hitherto been hidden by the more generalised description’. The completed finding aid was converted to Encoded Archival Description (EAD) and provided the structure of the web site, with hypertext links from the item to the digital objects. Additional material, such as Barton’s biography, the finding aid and some utilities are also included in the site.

The decision was made to digitise the material in-house—except for oversized (larger than A3 page) and three-dimensional items—due to the relatively small size of the project and the large number of interdependent procedural issues to be resolved. The project was seen as a learning process, presenting the Library with the opportunity to gather valuable information for the future projects. An A3 size scanner was chosen as most suitable for scanning of majority of items. A Conservation graduate with basic paper repair and handling experience was employed as scanning technician to digitise the items, and she was trained and supervised by the Library’s Imaging Services Branch. Imaging Services was also responsible for setting up the equipment and calibrating it to comply with the international standards for ICC colour management.

Initially, digitising of all three-dimensional and larger than A3 page items was going to be outsourced, but as the Library had already invested in high-end digital cameras for its’ digitisation program, these cameras were used to capture images, and the professional photographers from Imaging Services Branch were employed to operate them.

Digitisation standards were designed to achieve the highest quality possible at a reasonable price. These standards had been tested earlier during the Library’s Rare Maps Digitisation Trial project in 1999. Minimum spatial resolution for capture at 300dpi at original size was specified, and one of the important features required in the capturing devices was their ability to capture and process images at minimum 36-bit (RGB) level. For practical reasons (storage space, processing times) images were stored as 24-bit (8-bit per channel) RGB files. All images were captured and stored in colour, RGB mode. The storage format of choice for master images is TIFF. We chose not to use compression as it requires more processing power and introduces additional opportunities for file corruption. Copies of master images were automatically re-sized and saved as JPEG files for on-line delivery (thumbnails and viewing copies).

A large number of A3 size scanners were evaluated and we finally selected UMAX PowerLook 2100XL, which was relatively inexpensive but still suitable for the purpose of digitisation. The scanner’s maximum optical resolution is 800x1600dpi, with a dynamic range of 3.4D and internal processing done at 42-bit RGB. Two standard desktop computers with 128Mb of RAM and 17 inch monitors have been employed for scanning and processing images. As inexpensive and modest as they are, they proved satisfactory for our needs. Other existing equipment, like CD burners, monitor optimisers and Kodak Q60 (IT8) targets, were also employed during the digitisation stage. Initially, images were stored on CDs, as the server’s capacity was insufficient. The X-rite monitor optimiser and IT8 targets were employed to calibrate the
computers’ monitors to CIE D65 standard and to create ICC profiles for the monitors and scanner. To digitise oversized and three-dimensional objects Sinar Macroscan 23 digital back, mounted on Sinar large format studio camera has been used. Additionally, for some items Phase One PowerFX digital scanning back has been used. High-resolution lenses were used with each camera.

Software used in the process of digitisation comprised of UMAX scanning software, Adobe Photoshop for image manipulation and ICC colour profile application, and ColorShop for colour calibration. Additionally, CD burning and digital camera software have been used.

The identifiers used for items from the Barton collection—nla.ms-ms51—are also used as the persistent identifiers (PIs) online. This means that the Library has undertaken to continue to use the identifier unchanged in future to describe the item from its collection, including both the physical item and its digital surrogates. Persistent identifiers provide a reliable and stable way of naming items from the collection and these identifiers can be used in citations both online and in a print publication. Full PIs were physically included at the bottom of every JPEG image (except thumbnails).

The file naming system for Barton has 4 levels, reflecting Manuscript section’s folder hierarchy. The first level is a persistent identifier: nla.ms-ms51 The second level is a 4 digit decimal number reflecting series number. The third level is a 4 digit decimal number reflecting item number, and the fourth level is the letter “s” and a 3 digit decimal number describing the sequence number (reflecting page) followed by a letter to describe use (m = master, d = distribution, t = thumbnail).

The use of Persistent Identifiers and an associated resolver service are an attempt to solve the common problem of broken links that occur when resources on the web are moved to a new location. The approach taken by the Library provides both a stable and consistent naming scheme for collection items and a resolver service to redirect users to the current location of an item based on the persistent identifier. Having persistent identifiers and persistent web addresses for digital material will enable users of the Library's services to cite resources with confidence and build online virtual works that incorporate material from the Library's collection.

Metadata accompanying images have been organised in the form of Microsoft Excel spreadsheets and were mostly automatically derived from TIFF tags. The metadata is of the type required for management and preservation purposes. The full file name has been embedded in the TIFF tag 269 of every master image file.

As mentioned, during the first stage of digitisation images were stored on CD-Rs. Subsequently, as space has been made available, all images have been transferred to the Library’s server and, after creation of derivatives, master images were then moved from the server to archival storage tape. Local Area Network has been employed to move images from digitisation stations to the server. Though 100bps switches were used it was the slowest part of the digitisation process.
The Library already had permission to make copies of the Barton Papers available to researchers, however, a web site is a form of publication and under the terms of the Copyright Act the Library was bound to seek permission from all the copyright owners for the digitisation project. As it was impossible to contact them all 400 of them, the Library approached the Barton and Deakin families individually, and a notice of intention to publish was placed in the Government Gazette for all the other correspondents. The Library has not received any protests from the copyright owners.

Many sections of the Library have been involved in creating a digital presence for the Barton papers. Manuscript Section worked hard on the collection to make it ready for digitisation and to complete the finding aid. Preservation staff were involved in the project right from the beginning, making sure that our scanning procedures are safe and preparing material for capture. Information Technology Section struggled to provide the Barton project with enough storage space. Web Services and Multimedia were responsible for creating the web site and working out the details to make sure that everything works properly. Imaging Services Branch lent its expertise to install and calibrate the scanning system, and to train and supervise the scanning technician. They were also responsible for capturing oversized and three-dimensional items.

Making the project a reality was by no means an easy task. Right from the beginning the Library encountered many problems. Indexing the items and correcting filing and name errors, created additional work for the Manuscript staff. Missing items needed to be found and scanned long after the main scanning process has been finished. Initial problems with the storage space, and problems associated with moving large files across the LAN slowed digitisation down. One of the problems during the digitisation stage was trying to automate cropping of master images for the web as software employed to perform this task proved to be unreliable. In addition, the metadata needed refinement and the finding aid required re-mapping to make it work as the structure for the web site. The logistics of creating navigation tools was also not an easy task.

The Barton project was a learning process. It provided the Library with an invaluable insight into practical issues associated with digitisation projects. A less ambitious choice of collection would probably have made the task less painful. More detailed planning at the very beginning of the project would have avoided the re-scanning of some material. One of the many important lessons learned was the realisation that scanning is a very tedious and exhausting task. In order to maintain productivity, more scanning technicians should be employed for future projects and work shorter shifts. Another important lesson was finding out that we miscalculated storage space required to accommodate Barton images. We did not take into account the large number of temporary files that were created during digitisation process. Also, transferring finding aid to the web proved to be a very complex task.

Despite the difficulties, the Barton project was a success. It has provided the Library with the tools and solutions to deal with future projects and has indicated areas that need
improvement. The project produced a beautiful and attractive web site that will be of continuing interest to researchers and general public.