A virtual reading room: access to digital documents

Margaret Kenna
Director
Accessibility Development
National Archives of Australia

Introduction
The issue that challenges the National Archives today is how to promote wider access to and use of its vast and wonderful collection. The challenge is complicated by many factors including:

- the sheer size of the collection – about 300 kilometres of records;
- the value and unique nature and in some cases the fragility of the records; and
- a wide geographical distribution of Australia’s population which precludes direct access to the collection for many.

This paper initially focuses on the problems associated with making equitable accessibility a reality for all users of the Archives. It then discusses the Archives’ investigation into the feasibility of digitising records to increase accessibility, the steps taken to implement a digital service and finally the alliances that have emerged through this digital initiative.

Recognising the problem
I would like to set the scene for this section of the paper by presenting a case study on an average National Archives researcher and his research experience.

John lives 300 miles north of Perth. He is an enthusiastic archival researcher and uses the National Archives website and RecordSearch to identify material that could be useful for his research. After identifying records that could be of interest according to the title of the record, he emails the National Archives reference service asking for a quote for photocopying the records that he has identified. These records are located in Canberra.

The request is received in Canberra and an Archives staff member withdraws the identified records from storage, calculates the cost of photocopying the records, contacts John and advises him of the costs involved. John then has to consider these costs in relation to the likelihood that the records contain the information that he needs. He realises that he cannot afford to pay for copies of all the records so finally he selects those that have the most relevant titles.
The Archives receive John’s final selection and payment. The payment is receipted and the records are photocopied and posted to John. Another happy customer – or is he?

John received his package of photocopies but to his dismay, the records do not contain information of relevance to his research.

This scenario may be repeated many times each month across the country – we have not been able to gather statistics about failed research. The question is: How could John’s research have been better managed? In a traditional archival research environment there were other options. John could have travelled to Canberra to view the records or he could have engaged the assistance of a search agent to examine the records for him. Both of these options were costly alternatives to the photocopying decision.

Perhaps a more relevant question is: Why should John be placed at a disadvantage to people living in Canberra who can access these records and at no cost? This inequity has been something that the Archives could not adequately address in a traditional reference service environment.

Technology has been a most supportive partner to the Archives. It has enabled us to provide access to information, publications, standards and policies via the Internet, to anyone who needs it, regardless of where they live or work. Importantly for those who need access to archival material, it has provided, through RecordSearch, our collection database on the website, the capacity to access information about the structure of the Commonwealth Government and descriptions of the record groups that are created to support the functioning of the government. RecordSearch has given our dispersed client base – people like John the researcher in the case study – the ability to identify records that may be relevant to their research through a keyword search. However, there remained for users of the Archives the inability to fulfil their informational needs entirely, because they could not access the actual record online.

**Investigating the problem**

The National Archives could not ignore the issues the expansion of Internet access raised. As users increasingly expected to meet most of their information needs online and at the place and time of their choosing, the Archives had to either meet this demand or, over time, become increasingly irrelevant.

In 2000 the Archives initiated a project to investigate a wide variety of different forms of digitisation and web access. The starting point for the study was to provide access to our collection to remote users which, as far as possible, replicated the reading room experience. To do this, speed and
minimisation of cost for image capture and adequate legibility of the finished product were the essential criteria.

In many ways the imperative of speed which ultimately impacts of minimal cost of image capture related to the size of the Archives’ collection. The sheer size of the collection provides a potential for large volumes of requests to be lodged by researchers and the access to be requested to any record from the collection. I noted earlier in the paper that the collection comprises about 300 kilometres of records. To gain an appreciation of this vast collection you could think about these records, stacked side by side along the highway stretching from Sydney to Canberra. You would need to drive for three hours to pass by them.

The strategy for the project was to trial various digitising methods. Overhead scanners, digital cameras, book-eye scanners and flatbed scanners were tested. Microfilm to digital conversions were also tested. A variety of record formats were used to test each method’s ability to capture all types of folios. We were aware that World War 1 service dossiers presented a particular problem in photocopying, given the variety of ink and pencil colours, handwritten and typed entries, stamps and background colour on the forms. Therefore a selection of service dossiers was an obvious choice for inclusion in the trials because if digitising these difficult formats was successful, other less difficult records would be relatively easy.

An important facet of the project was the development of a cost-effective and suitable method for making the digitised images available through the Archives website.

Early in the project it became clear that all images should be attached to the item entry on RecordSearch on the website. This was:
- to remove any need to create of replicate metadata – a significant factor in keeping capture costs low;
- to maximise the chance of the long-term preservation of the digital data;
- to avoid the duplication of digitised images on the website;
- to maximise researchers’ chances of knowing of the existence of images of records which they might wish to consult; and
- because images on RecordSearch could be linked to other parts of the website if required.

**Strategies emerging from investigations**
The clear conclusion of the trials was that digitising by overhead digital camera met the image capture criteria better than the alternatives examined. This method provides adequate legibility with the greatest speed of throughput. It is also the simplest system for capturing images and the one...
that best replicates the reading room experience both in record capture and in
the appearance of the images.

This direction is clearly at odds with much of the literature about digitising
archival records. There are two reasons for this. First, most of the experience
leading to the formulation of the high-resolution, best image model has been
based on pictorial material or rare and attractive items, a significant
proportion of which will be likely to be required for illustrative and
publication purposes as well as for research. The literature, and the
digitisation discourse in Australia, also stresses the digital preservation of
objects.

Our project, however, concentrated on the file-based records commonly
requested in reading rooms, which are unlikely to have any significant use
other than for research. In this case, paying again to copy to a higher standard
any item that is required for other purposes in the future will result in a
substantially lower overall cost than capturing all images at a higher standard
‘just in case’. Second, because overhead scanning techniques markedly
increase the speed and reduce the cost of image capture, a significantly
greater number of images can be captured and made available.

Costs

The primary reason for recommending a system which uses digital cameras
for image capture and aims for minimal standards was that speed of
throughput is seen to be as vital as cost.

The marginal cost per image for the preferred option, based almost entirely on
the labour cost, depends on the number of images operators can capture.
During the trials the average number of images per minute for image capture
and processing was 3.26. At this rate, and at the APS2 salary level of the
operator, the marginal cost per image was between $0.11 and $0.14 depending
on the type of material being processed, with an average of $0.13. The cost of
equipment and the cost of developing the software added an extra $0.01 per
image per year. This is less per image than any alternative method examined
or known.

1 The following were considered in relation to the digital project and trials:
Anne R Kenney and Oya Y Reiger, Moving Theory into Practice: Digital Imaging for Libraries and Archives, Research
National Archives and Records Administration, NARA Guidelines for Digitising Archival Materials for Electronic Access,
Anne Kenney and Stephen Chapman, Digital Imaging for Libraries and Archives, Cornell University Library, Ithaca, NY,
1996.
Committee to the Commission on Preservation and Access, Commission on Preservation and Access, Washington,
DC, 1990.
Colorado Digitization Project, General Guidelines for Scanning, downloaded 31 July 2000 from
http://coloradodigital.coalliance.org/scanning.html
Linda Sorenson Colet, “Planning an Imaging Project”, in Digital Library Federation, Guides to Quality in Visual
The Digital Toolbox (http://coloradodigital.coalliance.org/toolbox.html) gives links to a wide selection of these;
Kenney and Reiger, p. 9 and in footnotes throughout, also provide a plethora of references.
However, we have found that further cost gains can be made by separating the image capture task from the image processing and loading task and using different classification levels of staff for the differing levels of capability required for each task. It was estimated that casuals working relatively short shifts could capture more images per minute than the 3.26 obtained during the trial. At four images per minute, the marginal cost for image capture (including the processing and loading time) would reduce to between $0.09 to $0.12 per image with an average of $0.11. At five images per minute it would be from $0.07 to $0.10 with an average of $0.09 and at six images per minute from $0.06 to $0.08 with an average of $0.07.  

The marginal costs per image do not, of course, reflect the total cost of making the images and loading them onto RecordSearch. The full costs include record retrieval and return to storage (calculated at approximately $0.09 per image) and storage of the digital images (calculated at $0.02). However, our contention is that these costs are incurred irrespective of the image capture system employed. The record retrieval and return to storage costs already apply to each folio of every paper record retrieved for remote user photocopying or when records are made available in National Archives reading rooms. In the digitising process this cost would normally be incurred only once – when the digital request is actioned. Therefore the philosophy is that digital capture once equals many uses.

Recommendations
The recommendations coming from this project were that:
- a digitising program be established based on digital capture by digital overhead camera;
- an on-demand service be initiated that allows users of the Archives to request that a digital copy of their selected record be loaded onto RecordSearch on the website; and
- high use record series be identified, digitised and loaded onto RecordSearch.

What is involved in implementing a digital service
The process of creating digital copies for RecordSearch has three components – capture, ImageStore and ImageLoader.

Capture
Capturing the digital image is very simple for the operator. It requires the operator to:
- be aware of preservation considerations at every step of the process;
- log the file barcode;
- place the file under the camera, aligned in a pre-set position;

---

• capture the image by depressing the camera shutter; and
• turn to the next folio and capture in the same manner until the whole file is digitised.

The guiding rules are to digitise from the top of the file down and to avoid dismantling the file unless it is necessary for legibility.

Hardware that is required for the process is:
• a digital camera which is attached to an adjustable stand for the overhead camera alignment; and
• a computer for uploading the captured digital images.

Software consists of two inhouse-designed programs, ImageStore and ImageLoader.

**ImageStore** rotates, crops, sharpens and adds contrast to the captured image without human intervention. It allows an on-screen review of pages copied, and at this point single pages can be replaced or redone if necessary. This program saves a large and a small copy of each raw image that has been produced during the capture stage. The small image is the default image that is loaded for viewing. Users of RecordSearch may select the larger image for print purposes. Images are captured at a resolution of 180dpi with average image file sizes of 129KB for large, 104KB for raw and 52KB for small images. Download time for the small default image – the read image – is short.

**ImageLoader** is the conduit for loading the digital images onto RecordSearch. This program will also load images that have been captured in processes other than the digital camera/ ImageStore mechanisms. It has the facility to replace and delete pages or whole files.

**Availability of the service**

On 11 April 2001, the on-demand service, whereby users could request digital copies of records to be loaded onto RecordSearch, began operation. The Archives did not publicise this new service as we could not predict how the processes that were trialed in an artificial trial environment would translate to an actual service environment. Nor did we have an appreciation of the volume of requests that would need to be handled by a service centre which was very much in establishment mode.

Prior to the service launch we decided that initially the service would be offered only for record items located in Canberra. We considered that this would give us time to refine processes, gauge the request volume and establish the correct infrastructure that was needed to provide a truly responsive service to users of the records. When the service is fully functional it will be extended to State Office holdings.
**Productivity**

Capture rates reached in the early period of operation have been averaging 1100 pages per operator per four hour shift. Each shift includes a 20 minute break which is taken through the four hours in 5, 10 and 5 minute breaks. The average capture rate is easily achievable for records in regular formats (ie where no dismantling of files, removal of pins, plastic sleeves, unfolding of maps etc is required). The capture rate can easily fall to as little as 2 to 2.5 pages per minute if this sort of manual preparation is required throughout a file.

Image processing can run at 4400 pages per operator per four hour shift. This is based on a processing rate of 20 images per minute. In practice, the processing rate is constrained by the rate of capture.

Our early experience shows a larger than anticipated number of files have required dismantling or special preparation prior to capture. This reflects part of the ‘unknown’ facet of operating in an actual work environment as opposed to the trial project environment.

However, we are confident that productivity rates can improve, in light of the possibility of further developments in the digital camera operation speed enabling another dimension of automation for the capture/store process.

**Requesting online digital records**

To request an online digital copy be loaded onto RecordSearch, users select one of the two buttons that appear on the item description screen. The buttons prompt requests for either a photocopy or an online digital copy of the item. The prompt for requesting a digital copy appears on the item information screen. Note that the item must have an access status of ‘open’ or ‘open with exception’ and at this point the item must be located in Canberra.
Example of Item Screen

A request for a digital copy is submitted via an email form. A user needs to have Internet mail enabled on their mail software in order to submit the online request. The user receives an electronic acknowledgment on submitting the request. The message notes that the request has been received and is being processed. It notes that the digital image will be loaded onto RecordSearch within 30 days.
Example of request screen
The digital copy online
When a digital copy is available for viewing online an icon appears on the item description screen.

Example of digital icon

![Example of digital icon](image-url)
Viewing the online digital copy

World War I - Lieutenant General Monash - Service Record - Page 1

Note the navigation tool at the top of the page. You can advance through the file, page by page, or jump to the page you require by entering a number in the ‘Page’ field. You will also note that there are version selection buttons at the top left side of the screen. The default brings up the ‘small’ digital image (eg 52KB) which is adequate for on-screen reading. However, we have provided the option of selecting the ‘large’ digital image for printing purposes. In practice we have found that the ‘small’ image usually provides a very legible printed copy.
Digital alliances – extending accessibility

During the past year we have worked with a number of bodies to explore the concept of increased accessibility and to make it a reality. The digital system that we have produced allows external sites to link to digital items in RecordSearch. This has a multiplier effect for accessibility in that some of the users who come to RecordSearch from other sites may not have had access to these records if it had not been for the link provided from their original search site. We need to continue to recognise potential alliances that will afford greater accessibility to our nation’s archives.

John Curtin Prime Ministerial Library

Commonwealth documents form part of the online digital archive that is presented through the John Curtin Prime Ministerial Library (JCPML) website. The National Archives and the JPCML have worked together in making these digital images available. The last group of Commonwealth files that have been made available on the Library’s website are provided through a link to RecordSearch on the Archives’ website. Digital capture once results in many uses.
University of Newcastle

During the project trial period, we were approached by Dr Wayne Reynolds who lectures in Foreign Relations at the University of Newcastle. He was interested in making digital copies of archival documents available to his students for their research course work. A number of foreign relations files identified by Dr Reynolds, were digitised during the digitising trials and have subsequently been loaded onto RecordSearch and made available online for his students and for any one who has an interest in foreign relations. This group of files, which covers aspects of foreign relations in Japan, Indonesia, Portuguese Timor and China, can be found under a Foreign Relations button on the website or can be accessed through RecordSearch. There is only one digital copy of each file but each can be accessed through two different points on the website.

National Centre for Hellenic Studies and Research - La Trobe University

An alliance with La Trobe University, which is at a preliminary stage, grew from the Hellenic Studies Centre’s need to gather together records that document Greek migration and other aspects of life in Australia for the Hellenic people. Rather than requesting photocopies of the relevant records, the Centre will identify the relevant records and the Archives will digitise and load them to RecordSearch. The Centre will then provide links from their online collection to the records onto RecordSearch. This will make a significant group of records available through both the Archives and the Hellenic Centre.

Government agencies

Digital accessibility is just as important for government agencies as it is for public researchers. There are a number of examples where an agency needs access to record series to enable their ongoing work even after these records are more than 30 years old. The difficulty in the past has been the need for these records to be in archival custody and publicly accessible (both legal requirements), while still available for reference by agency staff. This has involved, in some cases, the need for agency staff to consult these records in the Archives reading rooms or to have photocopies made for them. Digitisation will facilitate agencies’ remote access to the collection.

Office of Australian War Graves

In the case of the Office of Australian War Graves (OAWG), the agency had a microfilm copy set of grave registration sheets made to enable them to continue with their core work following the transfer of the grave registration sheets to archival custody. OAWG has worked with the microfilm for a
number of years but in recent months they have realised that the microfilm copy is deteriorating badly and recopying will be required. Fortunately they now have a new copying option available in the form of digitising. The Archives is working with OAWG to come up with a joint approach that is expected to result in the sheets being available on RecordSearch. This will fill the OAWG need for access to the information contained in the cards and also make these archival documents available for anyone who wants or needs to access them.

**World War II service dossiers**

Service dossiers, especially World War I and World War II dossiers are the records most highly sought by public researchers. For many years the service agencies’ continual need to access these records for entitlement information has overridden the requirement to transfer them to archival custody. In 1993 the WW I army dossiers were transferred to the Archives as the agency’s need for entitlement information decreased. As public interest in the WW I records increased, so too did the interest in WW II records. Public requests for access to these records has meant that there has been an ongoing and increasing need to temporarily transfer service dossiers between the agencies and the Archives.

It is expected that in the near future the service agencies will begin transferring the WW II dossiers to archival custody. This does not mean that the agency does not still have a residual need to access the records for entitlement information, but there is now the capacity, through digital copying, to make these records available for both public and agency access when needs arise. The service agencies and the Archives are working together to ensure that the access needs of all parties will be accommodated. This cooperative venture would not have been possible without the digitising program that we have initiated.

**The future**

Over the past five years we have witnessed the changes that technology has wrought in many facets of people’s lives. Who would have predicted that the Internet would become such a central part of our communication, business and entertainment? The Australian Bureau of Statistics (ABS) has documented the take-up rate for this technology. Figures in 1997 show 7.5% of households had access to Internet. The following year there was an increase to 19%, followed by 25% in 1999 and 37% in 2000. To give a more pertinent reading, in the 12 months leading to November 2000, 50% of all adults in Australia accessed the Internet. The November 2001 release of ABS statistics is expected to provide proof of further escalation of Internet use.

---

3 Australian Bureau of Statistics, 0146.0 ‘Take-up rate for modem and Internet use low’, media release 0146.0, 14 November 1997 and 0147.0 Use of the Internet by Householders, Australia: Summary of Findings, media release 0147.0, 16 February 2001.
In 1995 the Archives grasped the opportunity the Internet provided to make our services and research tools more widely accessible. It was this technological foundation that enabled the transition to the delivery of online digital records in 2001.

If we are to continue to provide accessibility and services that are relevant to our ever-changing environment, we cannot afford to ignore new technologies or the wants and needs of our users. There are always new audiences and new media to conquer.