

# All things to all people: Digitisation and research communities

Paper for Australasian Libraries in the Public Sector 2009 Conference: Interaction and Innovation

Daniel Wilksch  
Manager Online Projects  
Public Record Office Victoria

## **Abstract**

Public Record Office Victoria uses digitisation as a tool to assist and encourage researchers using Victoria's public records. As well as in-house digitisation projects, PROV has entered into a long-term arrangement with the Genealogical Society of Utah to digitise Wills, Probate and Inquest records. PROV is looking to increase the role digitisation plays in its Reading Rooms and is developing guidelines for the digitisation of records by Victorian public agencies. This paper explores the ways in which PROV's digitised material has enabled collaborative work among a variety of PROV's research communities.

## **Introduction**

It is possible that digitisation projects are becoming a bit quaint these days. We live in the world of Web 2.0. The internet is no longer a platform for the delivery of bulk content, it is now the source (and subject) of a billion contacts and conversations.

It's probably more accurate to say that digitisation has become a normal part of our society. The technology has matured immensely in the last few years. Digital imaging is ubiquitous, scanners are now built into the cheapest office machines as a matter of course, mobile phones record video. Storage is extremely cheap, or free, if you count the multitude of file-sharing, social, and other websites.

Meanwhile Google has embarked on the enormous project to make the contents of as many books as it can searchable online.

PROV has, like many other cultural institutions we have spent considerable energy over the last few years developing an internet presence.

## **In-house digitisation**

First, some history. PROV started digitising in 2002 what were at the time our most visible and high-value records, the collection of documents relating to the chase, trial and capture of the Kelly gang. There are nineteen boxes in this collection, mainly of documents culled from other record series because of their topical interest. It took us at least six months to get these scanned, and we abstracted around eighty documents from the total to publish in an online exhibition; imaginatively titled 'ned online'.<sup>1</sup> While PROV has had a web presence since 1996, this was our first experiment in publishing digitised content to the site. Indeed, you can do an archaeological trawl of our site to discover the web fashions of the day.

The eighty documents for 'ned online' were transcribed, and a story was written around them. We hired a web design company to come up with a flashy-looking site, and launched it with a seminar at the Old Melbourne Gaol.

---

<sup>1</sup> <http://www.prov.vic.gov.au/nedonline>

The website was the focus of that project and the rest of the documents digitised from that collection sat around on CD ROMs until much later when we were able to do something more secure with them.

Since then we have acquired four middle-of-the-range flat-bed scanners for photographic material and fitted out a room as a photography studio with flash lamps connected to a digital SLR camera for capturing volumes. The camera also allows us to capture large maps and plans up to about A2 size. Anything larger requires the stitching together of images – a tedious process that does not always produce a good result in a reasonable time – so we have arrangements with third parties to digitise these.

As well as the Kelly documents, we have digitised (or have arranged for commercial suppliers to digitise) several other document collections. Our selection criteria have been records that are high-value and high-use in themselves, such as our collection of records relating to Koorie people; or records that are useful to progress research, such as the collection of patient registers for Victoria's mental hospitals that we have transferred from microform to digital files and put on our website.

Currently we are engaged in a project to digitise our prison register series. These are high-use records that contain photographs, descriptions and detailed histories of 'guests' of Melbourne's gaols. They are physically large and difficult to handle, and at some point in the past many were water-damaged, leading to conservation issues. In short there are a number of good justifications for digitisation.<sup>2</sup> We have the volumes conserved, and use the digital camera to capture individual pages. Because in this series, each page represents a named prisoner, we are able to supply the images to our team of volunteers, who transcribe prisoner names and numbers to form a database. We can then publish the volumes and database to the website.

### **Partnerships**

In 2005 we were approached by the Genealogical Society of Utah (GSU) and formed a partnership to copy our Wills, Probate and Inquest records, from the beginnings of the Colony of Victoria from the 1830s to 1925.

The GSU began the project using a digital camera, the first time they had used digital capture over microfilm in the southern hemisphere. Because they did not yet have a method of publishing the images they captured, part of the agreement was that PROV would publish the images to its website.

The partnership works in the following manner. PROV has provided a room, in which the GSU have installed two purpose-built digital cameras linked to computers. Software to automate the operation of the cameras as much as possible has been developed by the GSU. The camera operator sits in front of the copy table. The records are unpacked, fastenings removed and pages unfolded. The camera operator works through the documents, imaging front and back, but ignoring pages with no text on them. Once the records are imaged, a second volunteer records basic metadata. The images are saved to portable hard drives and forwarded to the US for processing in Salt Lake City. PROV also secures a copy of the images and metadata for processing into its own digital repository.

The size of the GSU team varies depending on the availability of volunteers, but consists

---

<sup>2</sup> An example can be seen by entering a surname in the online index:  
[http://proarchives.imagineering.com.au/index\\_search.asp?searchid=53](http://proarchives.imagineering.com.au/index_search.asp?searchid=53)

of a lead couple from the US and several local volunteers.

A second project has also been undertaken by the GSU in partnership with us and the Victorian Association of Family History Associations (VAFFHO), that of transcribing names and details from the register of probates that controls the files being imaged. These registers were also digitised, and pages to transcribe are delivered over the Internet to volunteers working in their own homes to transcribe. The software to display the images and capture the transcribed information has been developed by the GSU, and has meant that over two hundred and fifty volunteers have been able to take part in this project, both in Australia and internationally.

The partnership with the GSU has been very successful, with over 500,000 files to be copied over the course of the project and made available for free through our website. We are hoping to use it as a model or inspiration for other projects with other groups. We have been approached for instance by community historical groups in relation to rate records.

Another outcome of the project is a demonstration of how effective digital cameras are at imaging large amounts of material quickly and with minimal risk. While the process of capturing an image with a digital camera is not as intuitive as using a photocopier for instance, once a camera workstation is set up, work can go very quickly. The GSU teams routinely create 50,000 or more images a month.

### **Making images available**

One can split digitisation workflow into the processes of image capture, image management and image access. These processes may look very different depending on the reasons why the image is being captured and the uses to which the image is intended to be put. Managing a digitising project is sometimes about managing the process to meet differing intentions for the image.

The practical challenge is that a digital image may be an image of many things. It may be a photograph, or a negative. It may be a single page from a file or a bound volume. It may be a large plan. Each of these requires a different management regime. A system that manages each image individually will work well for collections of photographic images, but the pages of a digitised book stored and accessed using the same system will be much harder to manage. The types of metadata needed to reference each of these looks different.

I want to explore two factors that influence our management of digitised images.

The first factor is simply that we are an archive and use archival principles to look after large collections of data. Our philosophical approach is to document and expose the system of organisation of any large body of data, so the user of that data can use those systems to access particular pieces of information. In crude terms, we catalogue recordkeeping systems rather than the records themselves.

This approach is well-suited to managing unique collections of records, but has to be thought about carefully when used to manage the multiple copies of records that are a feature of the electronic age.

Thus for instance, our volunteers have spent over fifteen years transcribing the immigration records of Victoria onto firstly microfiche and latterly a computer database freely available through our website. This is another form of digitisation if you like. We

have something like two-and-a-half million names transcribed now and more are added every year.

Naturally we have to preserve this immense volume of data, and we have accessioned the results of the project into our catalogue. For example,<sup>3</sup> we have catalogued both the original passenger lists (split into British, Foreign and New Zealand Ports), and the microfiche copies of those lists. We have also catalogued the microfiche copies of indexes to the microfiched lists. This is less than perfectly straightforward.

Indeed, most site visitors never see this, because we instead direct them to the transcribed and searchable online version of the index to the microfiche lists, which is not catalogued.<sup>4</sup> This is an example of us recognising that many of our clients will not use the management system for the records but a simpler access system.

The second factor influencing our management of the digitised images is the regime we have developed to look after electronic records over the long term. This regime is called the Victorian Electronic Records Strategy, and consists of a set of standards and guidelines that are mandated for use by Victorian Government agencies, a file format called a VEO (or VERS-Encapsulated Object), that we wrap around digital files to authenticate the file and convey information about its recordkeeping context, and the Digital Archive, which is the hardware and software designed to hold VEOs securely. It has a capacity of 14 terabytes, expandable at need.

VERS is primarily intended to manage born-digital electronic files – the contents of email systems and Electronic Document and Records Management Systems – but in fact a significant portion of what we have been feeding into the system since it went live in 2006 has been digitised copies of physical documents, both from our own and other agencies digitising programs. The transfer of digitised records will be covered shortly, but first an explanation of the ways in which the digital files are accessed.

The VERS strategy mandates that records be transferred in open, widely-used formats to simplify the process of format management over the long term. Acceptable formats are TIFF, PDF, text files, the MPEG-4 format, JPEG and so on.

Our task is to produce a file that can be imported in to the Digital Archive and managed over the long term, but which is also useful for access in the short term, in keeping with the priorities of our digitising program.

Unlike some image management systems, all of the digitisation we do currently, except for photographic originals, produces one copy that serves both as preservation and as the access copy. This is possible because relatively cheap broadband access is becoming widespread, so we do not need to drastically reduce the quality of the digitised files to deliver them over the internet.

Returning to the digitised wills and probates produced by our GSU partners as an example, the digitised images are imported into our Digital Archive and managed through our archival control system. However, we associate the images with the physical originals

---

3 Search the PROV online catalogue, <http://www.access.prov.vic.gov.au/>, using the Simple Search screen and the keywords 'passenger AND list'.

4 <http://www.access.prov.vic.gov.au/public/PROVguides/PROVguide050/PROVguide050.jsp>. The circle will be complete if we begin a project to digitise the original lists or their microfiche copies.

they are copies of, and do not expose our recordkeeping information around the images to the researcher.

We also have the facility to publish catalogue shortcuts to the website. As mentioned before, the catalogue mainly works at the recordkeeping system level, rather than the item level. Where we have data at a lower level than this, we have the option to import it into an index, which we can point at the digitised image online.<sup>5</sup>

Instead the researcher does a name search (as mentioned, the transcribed index to the wills and probates is itself the result of several months work by volunteers over the internet), is directed to a catalogue page, and given the option of downloading the image or viewing the original in our North Melbourne Reading Room.<sup>6</sup>

it should also be noted that once the indexing of the Wills and Probates is complete, we will be importing name-level data into the catalogue, so researchers can either search the catalogue directly or use the index as a shortcut, depending on their needs and level of expertise.

Managing our photographic originals is the exception. For these we are producing originals of 200 megabytes in size in TIFF format, which are not suited to most people's expectation of how images are used over the internet. For these images we have two copies in the digital archive – a master copy in its own series which is not exposed to the public, and an access copy attached to the series for the physical original.<sup>7</sup> Photographic images are also available in several other ways – we are members of the Picture Australia initiative run by the National Library of Australia, we are starting to put up images on the 'Flickr' website,<sup>8</sup> and we have recently posted many of our images to a wiki that we have set up, together with an open invitation for those interested to work on captions and add any other information that they have.<sup>9</sup>

### **Digitising and records management**

One of the indicators that many organisations are in a transition period from paper records management systems to digital records management systems is that many requests for digital transfers are in fact to transfer digitised copies of paper material.

Another piece of work we are engaged on at the moment is the compiling of advice to government agencies about the capture and management of digital copies of public records. This is partly in response to several new pieces of legislation that have come out in Victoria, most notably the *Crimes (Document Destruction) Act 2006* and the new *Evidence Act 2008* passed but not yet in force. The new *Evidence Act* removes the 'original is the best evidence' rule, which means that there is no legal impediment to introducing a copy of a document. The weight that the court gives to that copy is, however, dependent on evidence of a process that authorises the copying. The advice we are preparing is intended to assist agencies with putting that process in place. Additionally, if that process is in place, and digitisation is carried out according to

---

<sup>5</sup> See <http://www.access.prov.vic.gov.au/public/PROVguides/PROVguide066/PROVguide066.jsp> for an explanation of PROV's archival control system.

<sup>6</sup> <http://www.prov.vic.gov.au/access/probate.asp>. See also our passenger lists online, mentioned previously, as an example of information within original records being exposed online.

<sup>7</sup> See

<http://www.access.prov.vic.gov.au/public/component/daPublicBaseContainer?component=daViewSeries&entityId=12800> for an example of this arrangement.

<sup>8</sup> <http://www.flickr.com/photos/public-record-office-victoria/>

<sup>9</sup> <http://wiki.prov.vic.gov.au/>

standards that we are developing, agencies will be able to authorise the destruction of originals that have been copied.

## **Conclusion**

I can make two overall points about the various projects and initiatives mentioned here.

The first point is that the relationship between the original and the digitised copy online can be quite complex. It is a common comment or worry that looking at something online is no substitute for seeing the original close-up. Archives are like museums in that they hold unique cultural artefacts and seek to provide the context of the creation of those artefacts. Archives are also like libraries in that they have a role authenticating and authorising information. In our case, we are ultimately expected to support the operations of the courts in providing valid evidence of decisions made and actions taken. I would contend that for our records – regarded as information, as opposed to being regarded as cultural artefacts – the online version is just as good, and a lot more convenient than the original. Indeed digitisation is the first step in moving that information out of the systems in which we manage it into systems in which it can be used in a variety of ways.

This leads to the second point, which is related to another expectation that we are currently grappling with – the extent to which we are expected to work like the Internet, or specifically the culture of search pioneered by Google. In some ways PROV operates in a mode similar to Google – we provide information, but not necessarily a reference service to help you locate the right information for your needs. We are discovering that digitisation can help us meet this latter expectation by opening up the possibilities for partnerships to develop the lower-level detailed description of records that our researchers are seeking. Digitisation and the internet mean that indexing projects no longer need to take fifteen years.

The original purpose and justification for digitising is still valid – the recent fires in Victoria and the collapse of the municipal Archives in Cologne Germany at the beginning of March demonstrate that one of our roles is to reduce the risk of losing information.

But more than this, the drivers of our projects have always been about communicating information to the public more quickly, and increasingly about opening up channels for them to communicate back.