

Heritage Map Digitisation: An adventure in applying aspects of Digital Preservation Policy

Peter McGrath
Digital Preservation Specialist
State Library of Victoria
pmcgrath@slv.vic.gov.au

Shirley Firth
Image Resources Manager
State Library of Victoria
sfirth@slv.vic.gov.au

Abstract:

This paper discusses some of the key principles of the State Library of Victoria's recently released Digital Preservation Policy and how these were applied to a heritage map digitisation project, the Melbourne and Metropolitan Board of Works Maps Digitising Project. Principles discussed include image capture standards, image metadata, storage and media. The project presented many challenges and twists and turns along the way- a surprising adventure for the project team.

In late 2003, the State Library of Victoria (SLV) appointed a dedicated Digital Preservation Specialist. As part of the SLV's ongoing commitment to the Library of the 21st Century (www.slv21.com), this represented a new direction for the Library. While preservation and conservation of analogue material including pictures, books, manuscripts and maps was well-established within the Library, as with most institutions in Australia, the framework and discipline of Digital Preservation was still in its infancy. With the large amount of data being created in a digital form, and even more being created from analogue originals, the task of preserving this data could not be ignored.

The Digital Preservation Policy

A key outcome of the position of Digital Preservation Specialist has been the development of a Digital Preservation Policy, which was released in April 2005. The Policy includes a number of principles and many of these have been detailed into procedures/guidelines that sit immediately below the Policy. The principles include:

1. Image Capture Standard

The Library's image capture standard is known as the A400 standard. This name was derived from a baseline resolution at A4 (metric document size) and captured at 400 pixels per inch. As of 2005 the Image Resources Group at SLV is capturing selected heritage collection material at this standard. There are specific cases where this standard must be interpreted, and also practical considerations may mean that a compromise has to be reached.

2. Master Tiff Files

As TIFF 6.0 is considered a de facto archival image format, it was selected for its stable properties and its ability to be read by most imaging software. All images that are captured from heritage collection material will have a master file (TIFF) which will then be repurposed to suit.

3. Image Metadata

All Library master images will have selected metadata, including fields such as capture source, and an association with its original analogue material via the catalogue record. This links the digital version to the original item via the bibliographic number. See Figure 3.

4. Self- Supporting formats

Proprietary formats will be avoided as far as practicable, with a preference for published file formats. TIFF 6.0 satisfies this criterion.

5. Storage

The underlying principle of storage is that the Library avoids the risk of relying on one source of data, (DVD or Hard Disk) and that we have a systematically backed up method of file storage.

6. Media

All media selected for any SLV project will be of archival quality will be tested to ensure compatibility with all hardware (readers and writers) at the Library. Various tests were carried out, and other institutions such as the National Library of Australia were consulted before we decided on a type of disk.

In mid 2004, the Library proposed to catalogue and investigate the feasibility of digitising a series of oversized plans from the SLV Collection. The Digital Preservation principles, particularly those relating to image capture and future proofing were pivotal in the decisions we made in the ensuing Melbourne and Metropolitan Board of Works (MMBW) Maps Digitising Project. At the onset of the project, we didn't realise the extent of the adventure we were about to embark upon.

MMBW Maps Digitising Project

The main objective of the MMBW Maps Digitising Project was to develop an understanding of the issues involved in the digitising and delivery of map material via the Voyager online catalogue, and to put into place processes for the ongoing digitising of map material at the Library. Other objectives included:

- To provide digital access for both onsite and offsite users to high-use historical Victorian maps.
- To reduce the number of visits to the Map Collection to view material.
- To preserve the fragile original MMBW maps by making a digital alternative available to users.
- To create a high-resolution master file that could be repurposed for other outputs than web delivery including a digital copying service.

Description of the plans

The MMBW maps series comprises approximately 4000 detailed maps (or plans) with a scale of 40 ft to the inch and very large in size (130cm x 80cm). They are amongst the highest used material in the State Library of Victoria's Map Collection. The plans are used by family and local historians, local government, students and researchers.

The series was produced over a thirty year period from the early 1890s to the 1920s (hence no copyright issues) and is the most comprehensive record of Melbourne's urban development held by the SLV. The Library holds only one set, and many are in a very fragile condition, particularly the maps showing central Melbourne streets.

Access to the plans is via a master index map and a series of 39 index plans (scale 1600 feet to 1 inch). In effect, there are three tiers: the master index map, the index plans and finally, the detail plans.

Project content

With funding from the SLV Foundation, we set up a Project team to catalogue and digitise the MMBW plans and to provide user access via the Voyager online catalogue. The Project team comprised 8 staff from a number of areas of the Library including the Maps Collection, Cataloguing, Technology Services and the Image Resources Group, the latter including the Digital Preservation Specialist. We recruited an XML programmer for the Project due to the large research and development component. We specifically wanted a programmer who would be an integral part of the project team, who would investigate and make recommendations regarding methodology, hardware and software to be used, and who would share expertise, write procedures and provide training to Project staff. One of the purposes of

the Project was to put into place procedures for the on-going digitising of map material at SLV. It was not just a matter of selecting appropriate software; we wanted to be able to continue with the maps digitisation program ourselves when the programmer's contract expired.

Because of the nature of the MMBW plans series (i.e. the three tiers) we wanted to link the various levels of plans both through the bibliographic records on the catalogue and, visually, through the digitised plans on the html pages. We wanted to create links up and down from index and detail plans, as well as links to adjoining plans at the same level.

We decided to catalogue the entire series of plans using basic cataloguing templates, but to digitise only the master index map, all 39 index plans and some 750-800 detail plans commencing with the CBD, the highest used material, and then radiating out to the suburbs. This was due partly to financial considerations but also to the fact that many of the plans for the outer suburbs have little content, consisting of large blank areas and there is little point in digitising these.

The XML programmer was charged with researching and developing a suitable mechanism to deliver the plans electronically via the catalogue, allowing users to zoom, pan and navigate through larger and smaller zoom levels. In addition to the zoom facility, we wanted the programmer to develop an interactive web interface that would allow the aforementioned linking capability between the three levels of the plans.

Applying the Digital Preservation Policy

The remainder of this paper will describe how the Digital Preservation Policy guided and influenced many aspects of the Project. The nature of the project and its many components meant that we needed to experiment, and test various technologies and techniques. While we had established principles of the Digital Preservation Policy, they were to be stretched and teased in the development of the Project. The adventure had begun...

Image Capture standards

The SLV's image capture standard sets out the resolution at which collection material should be captured. While it is sometimes preferable to go for the highest available resolution, with some collection material, the *informational content* is the biggest determining factor. Pictorial material typically would require higher resolution captures due to much more detail. For the MMBW plans, we decided that the smallest amount of detail in the originals (the lines of the drawings) was able to be captured by a 200ppi file. We decided to capture in RGB Colour Digital Files, and Adobe® RGB is the chosen colour space. This meant that our images would conform to the standard operating environment at the Library. They could be read, and were of a uniform format that we knew we could cope with well into the future.

We tested and trialled capture standards at length before embarking on the project. Initially we planned to do the scanning in-house using the Library's Fuji Xerox Docuwide 2050 wide format scanner. However, we found that it could not produce a good quality output suitable for our purposes. It was only able to produce the equivalent of a digital photocopy and it was only 1 bit. This unit, while suitable for many black and white jobs, could not reproduce the subtle details of the plans.

We then did some trials with our digital camera (Phase one H20). We even tried capturing the plans in four sections and stitching these quarters together digitally. It produced a reasonable result, but it was very labour intensive, especially given that we were faced with the task of dealing with up to 800 of these plans. After reviewing our outputs from our own hardware, we found that we did not have the right equipment on site to produce a satisfactory result, so we started looking at outsourcing the task.

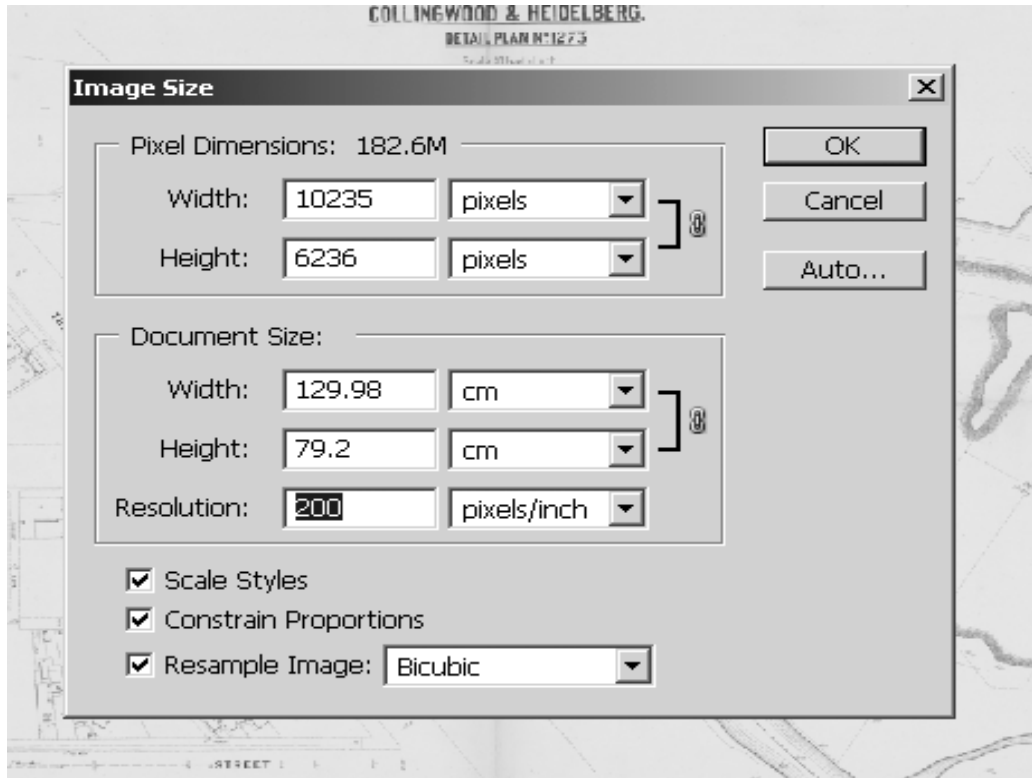


Figure 1. Plan Resolution

A small selection of plans, together with Technical specifications written by the Digital Preservation Specialist, was taken to three digital laboratories to get some sample scans. Colour tests were carried out at 200 (see figure 1), 300 and 400 ppi and the results from the three labs were quite varied. One lab did not appear to know how to handle fragile material and showed little interest in learning about the Project, whilst another (our chosen service provider) took a vested interest in the result, and produced excellent quality output. Quotes were also requested at this stage, and the prices received varied from approximately \$15 per scan through to \$60 for a very high resolution output.

It should be noted that a major concern for us with outsourcing was that the fragile collection material would leave the library, thereby putting it at risk in transit and in handling by outside service providers. To this end, Map Collection staff recommended that the plans be transported and scanned through purpose-made Mylar sleeves. At a later stage in the Project, visual checking led to the discovery of some minor Newton's rings. The phenomenon of Newton's rings is an interference pattern caused by the reflection of light between two surfaces – in this case between the scan head and the Mylar. Fortunately our service provider acquired a new scanner that eliminated the problems and we were able to continue using Mylar sleeves for the protection of the original plans without compromising the quality of the scans.

The costs of media needed to be considered also at this stage, as was the possibility of purchasing a portable hard drive to transport the files. We decided that DVD-R media was the best option, given that this storage media was relatively cheap and was able to carry a decent batch (25) of the images, the maximum weekly output of our chosen service provider. Given that the project had a finite budget, and we wanted to get more of the plans captured, the final 200ppi (c. 182MB per file) was chosen as the project standard. This resolution in this particular case was sufficient for both preservation and detail rendering purposes (see Figure 2). The plans were to be scanned at 100% scale, which was vital in maintaining the 40 feet to an inch scaling in the original plans.

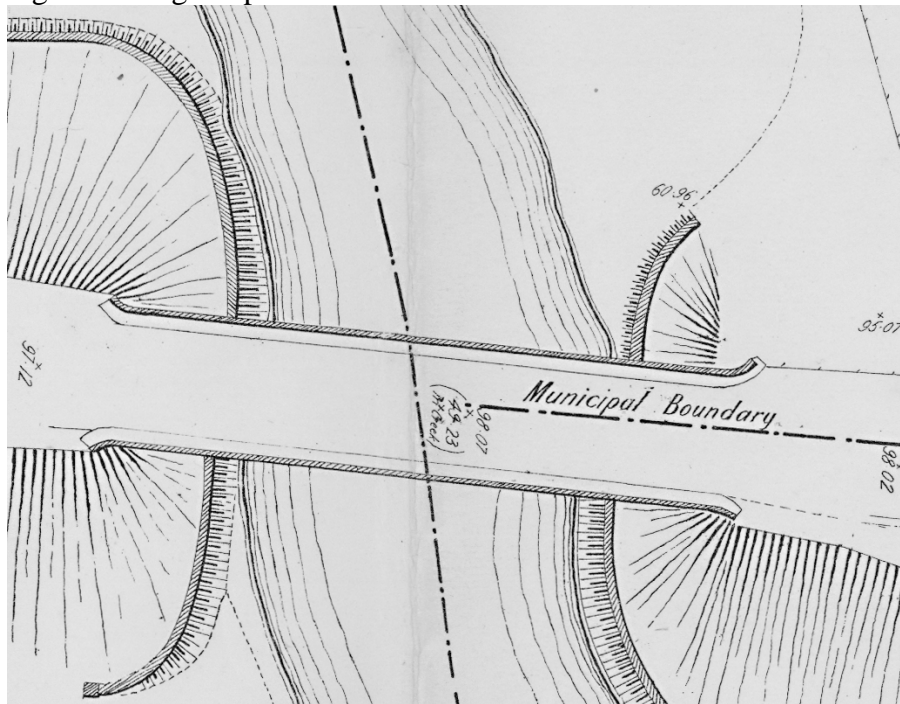


Figure 2. Detail of a bridge from plan 1273

Figure 2 shows the finest detail from a plan - fine pen strokes and lines are captured faithfully in the digital copies (original in 8 bit RGB Colour). The catalogue record with key words (contents/ summary) is shown below.

Corporate author: Melbourne and Metropolitan Board of Works
Title: Melbourne and Metropolitan Board of Works detail plan, 1273, Collingwood & Heidelberg [cartographic material].
Publisher/Date: [Melbourne] : MMBW, 1904
Description: 1 map ; 79 x 130 cm.
Map data: Scale 1:480.
Series: 1:480 scale plans (detail plans) 40 feet to 1 inch ;
MMBW map series. 40 feet to 1 inch. ;
Series no.: 1273.
Contents/Summary: Heidelberg Road -- Dwyer Street -- The Esplanade -- Spensley Street -- O'Grady Street
Subject(s): Sewerage Victoria Melbourne Maps.
Fairfield (Vic.) Maps
Clifton Hill (Vic.) Maps.

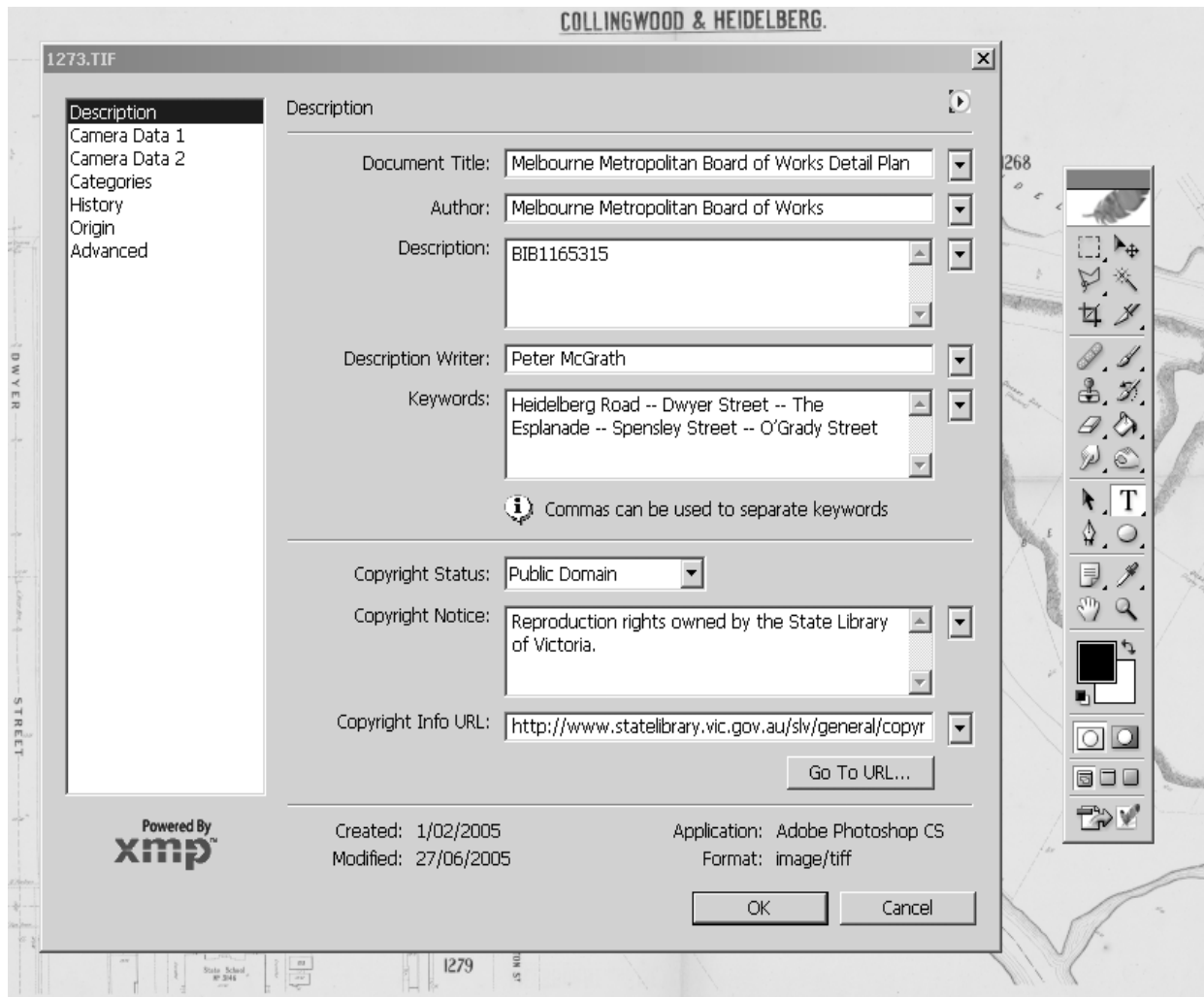


Figure 3. Metadata in Adobe® Photoshop® CS [1]

Metadata (embedded in Photoshop®)

When the images were scanned, we established that a set of metadata should be added to the image from the SLV catalogue (see figure 3 and text above it). This is part of our Digital Preservation Policy, and allows us to add value to our images, making them more usable. The key words (suburbs and streets) from the bibliographic record allow the images to be searched offline, before the image is linked to the catalogue.

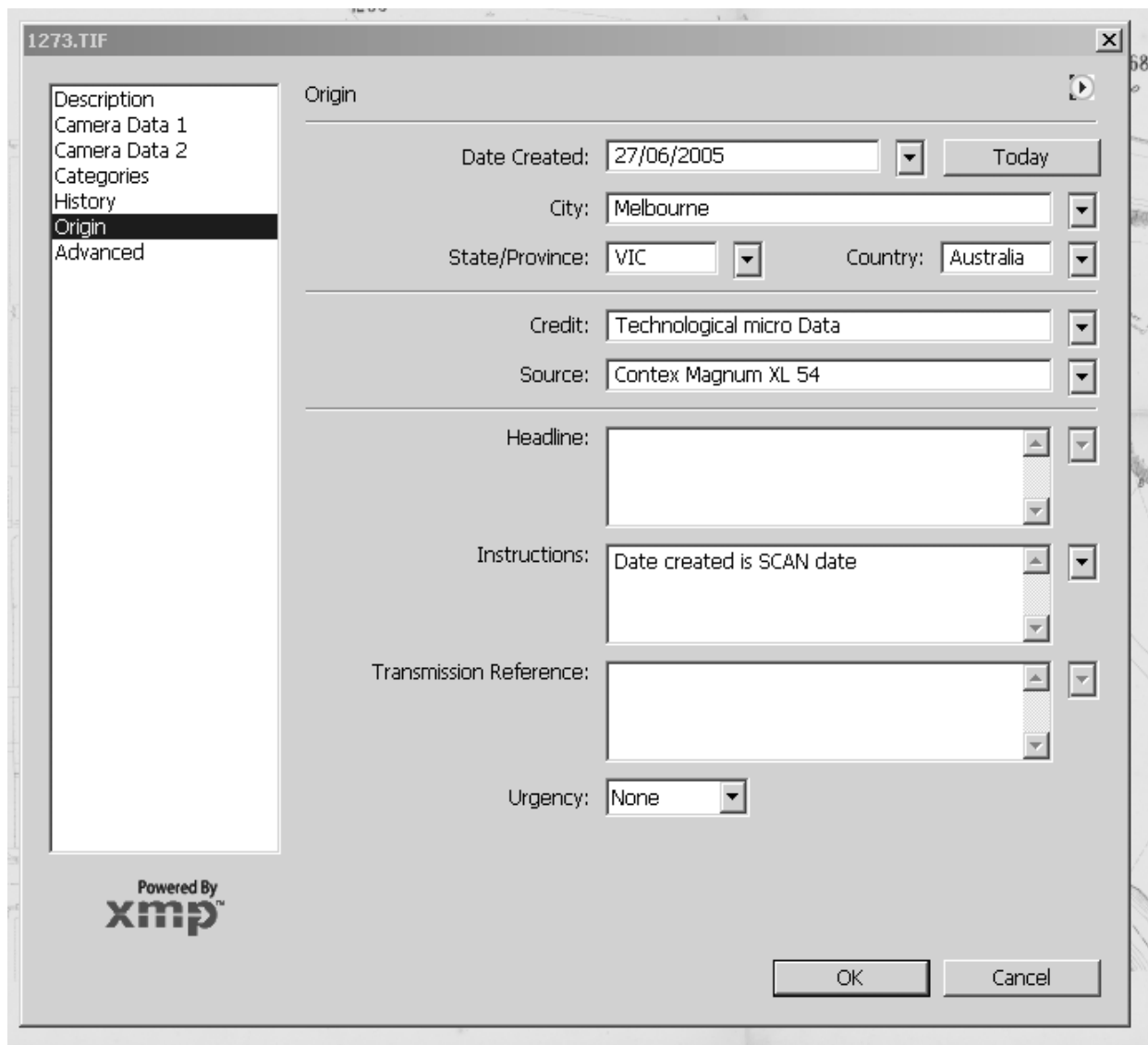


Figure 4. Metadata in Adobe® Photoshop® CS [2]

Figures 3 and 4 show the metadata screens in Adobe® Photoshop® CS (File Info). With our chosen service provider, we set up a procedure whereby we supplied a template to them, and at capture, each map was searched on the SLV's on-line catalogue, and various MARC fields were extracted and placed into fields in the File Info. (See Figure 4.) When the files were returned to the library (on DVD-R), we were then able to automatically extract this metadata into the Cumulus® program (see Figure 5). Cumulus® is a Digital Asset manager by Canto software. We compared several digital asset managers, and the Library is likely to invest in a more elaborate Digital Object Management System (DOMS) in the next few years that conforms with the global library environment. Initially, this software has been selected for images for the Image Resources Group, and has been a very useful tool. Its ability to read metadata from the file so it can be searched and sorted has been very useful.

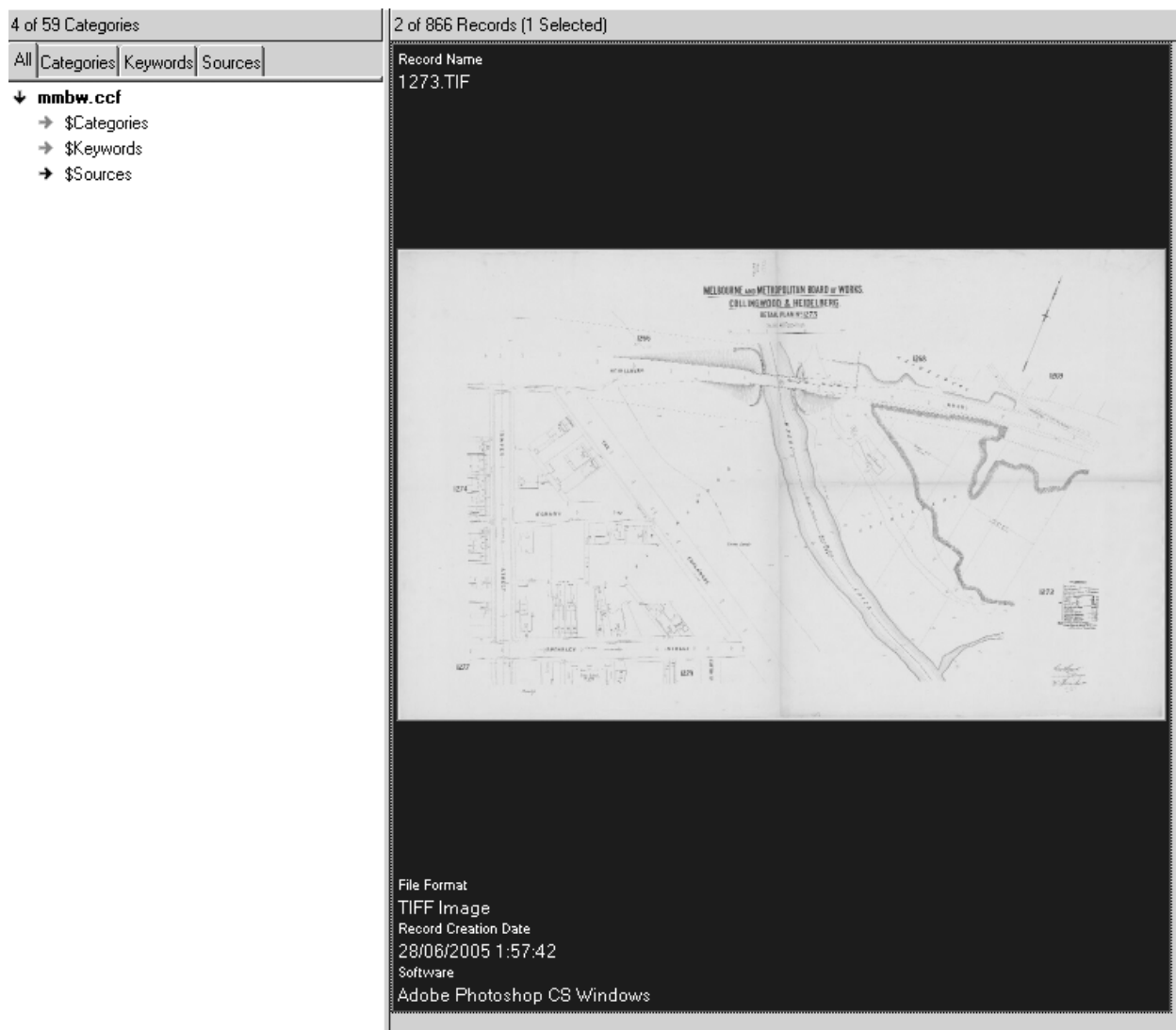


Figure 5. Cumulus® Browser window

This Cumulus® explorer window (Figure 5) shows how various file information is extracted from the file header (This can be customised). Thumbnails of each image are produced automatically when an asset is catalogued, and this allows an asset to be searched visually at first. Various fields from a wide range can be shown at the thumbnail view, allowing each asset to be compared by many parameters. In this view the File Name, date of creation and Software is shown. This view is set by the Cumulus administrator, and if needed it can be customised by each user depending on the level of permissions they have had assigned.

Further information on Cumulus® can be found at <http://www.canto.com/>, and the software version discussed in this paper is Cumulus® 6.6 – Archive Server.

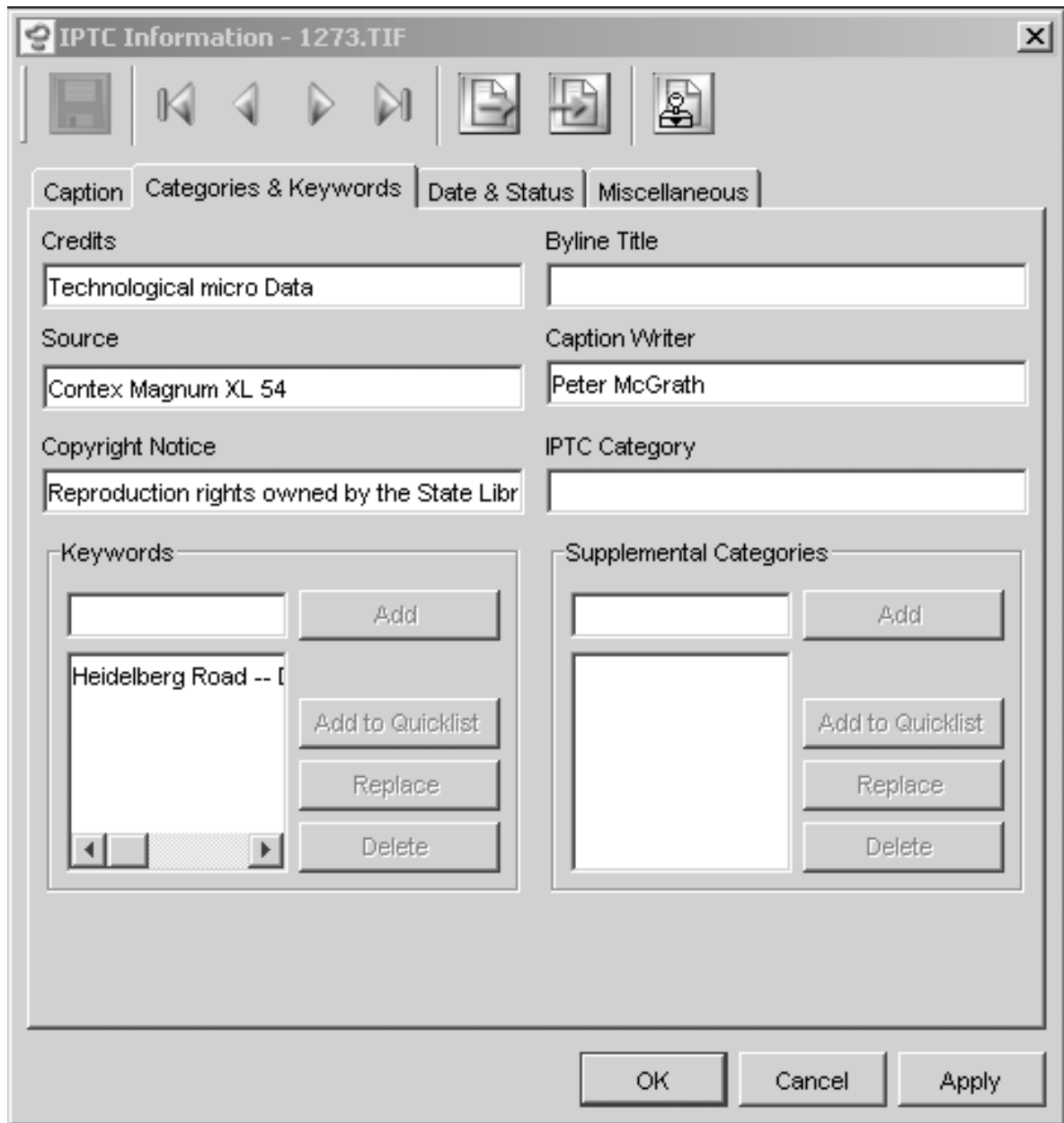


Figure 6. IPTC info in Cumulus® (extracted from the image file)

This metadata (figure 6) allows the archive of images (in Cumulus®) to be searched by street or suburb or other detail taken from our catalogue record. It also shows the scanner used in the creation of the image, something that may be of interest in years to come when comparing technologies and outputs. In terms of the workflow of image capture and metadata input, the best time for adding this metadata is at capture, especially given that DVD (read only) media was chosen.

Future Proofing of Digital Objects

We considered carefully the following elements of future proofing of digital objects as set out in the Digital Preservation Policy and underlying Procedures:

a) Self- Supporting formats

The selection of TIFF as the archival preservation format adheres to this part of the policy. As TIFF is a published standard (*see 2. Master TIFF files under Digital Preservation Policy*), we see it as a robust format for the master files from which all derivatives are made.

b) Storage

The files are initially copied to DVD-R (the most common DVD format to date), and then copied onto the SLV mass storage system. The SLV server is subject to our ICT policy in terms of backup and has a very low risk of loss. We also copy the files to a second DVD, which is kept in a fireproof safe as the master. The original DVD is used as the working copy and stored locally for easy access. DVD is the primary data source, but it must be considered as *part of* the whole process of data storage and management.

c) Media

We carried out tests on various DVD-R brands as to their stability and error rates, and we selected the best archival DVD. Although screen printing (on the surface of the DVD) was offered by the service provider, we rejected this as it was a variable that could increase the risk of data loss in the printing process. Some inkjet printers have interfered with the data layer of the disk, and adhesive stickers present more problems as well, in that some of the stickers have an acidic base and eventually eat through some of the data layer. Instead of these options we decided to hand write on the DVDs with an appropriate alcohol-free marker.

In the case of collection items stored on disk, the carrier (DVD, CD, Floppy Disk, Tape etc) is in principle to be treated as a temporary medium, especially given the relative fragility and instability of such data carriers.

d) Management of files

As illustrated in Figures 5 and 6, we have chosen Cumulus® as our Digital Asset Manager, and this allows us to search and retrieve our master files easily. We have looked into a number of products and shared experiences with other institutions. Cumulus has allowed us to share our work easily (at this stage we can have 3 concurrent users) and the data is exportable as XML should we decide to move to another system later. We have created a catalogue of images in this program, allowing us to track, search and retrieve the state and details of all images (online or on DVD).

Project outcomes

As at December 2005, we have catalogued all 38 index plans and all 4000 detail plans, with links between the various bibliographic records incorporated. We have a collection of over 650 high quality TIFFs with embedded metadata, stored appropriately on the Library server and managed by a Digital Asset Manager. These master files will be repurposed for other outputs than web delivery including a digital copying service.

Our programmer investigated in-depth a number of wavelet-based file formats, such as MrSid, Zoomify and JPEG2000, in order to deliver a viable web surrogate with sufficient clarity and functionality to meet the standards required. As a result, JPEG2000 has been utilised to create the required interactive web interface with a zoom facility and a linking capability between the three levels of the plans.

JPEG2000 (JP2) was selected as it has royalty free patents; it has been formulated by the Joint Photographic Experts Group (JPEG) which made the very successful JPG. This file format facilitates easier cross platform solutions and browser integration. We used the Kakadu® software for the creation of JPEG2000 files from the Master TIFFs. Kakadu® is an optimised, small memory footprint, C++ implementation of the JPEG2000 standard. It is written by Dr Taubman, Associate Professor at the University of New South Wales and the principal author of the JPEG2000 Verification Model software.

We are now able to provide web access to high-use fragile MMBW plans, preserving the originals and providing a digital copying service to our users, thus meeting our objectives. Figure 7 shows the Web interface- a breakthrough in image display for the library.

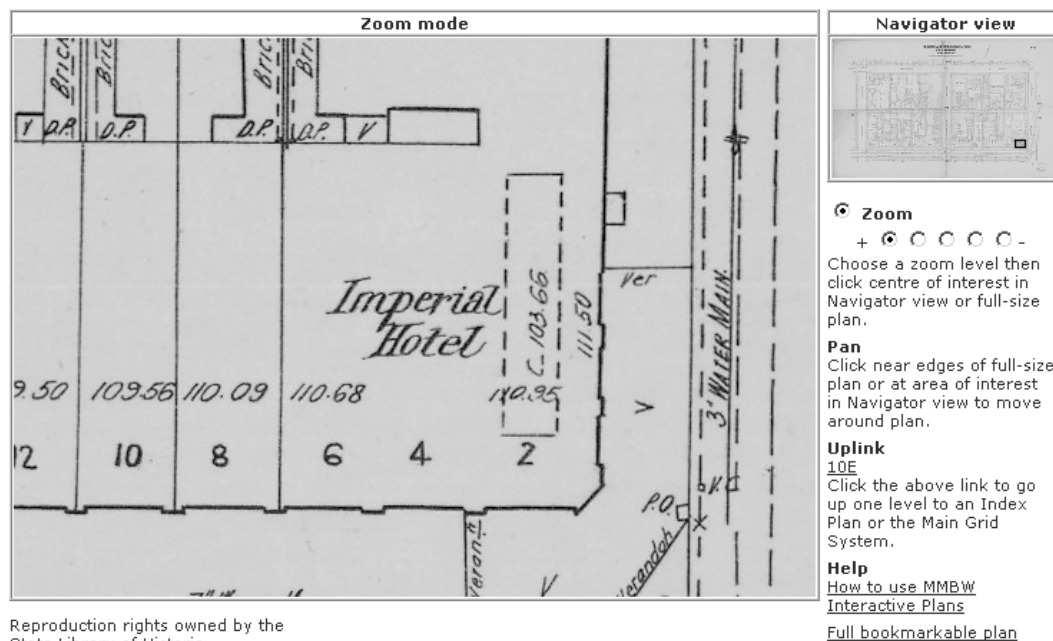


Figure 7. SLV web interface with JPEG2000 image

Conclusion

The State Library of Victoria's Digital Preservation Policy provides guidance and authorisation on the preservation of digital materials and ensures the reliability and long-term accessibility to them. The Policy forms the pillar of a program for digital preservation. It has proved itself to be practicable and flexible – it is being followed by many staff to date. As a result of this project, we now have a procedure in place for every aspect of the digitising and web display of maps. These can now be applied to other documents like drawings and architectural plans.

We have learned many things from this project including:

- The use of JPEG 2000 and how to implement and automate this process.
- Valuable lessons in how to negotiate and establish a great working relationship with service providers.
- Good internal and external working relationships - both with vendor and cross-divisional members of the team.
- Metadata and techniques for using catalogue and key words, and the value adding process.
- How to use Cumulus as a searching and data management tool.

By applying the relevant principles of the Digital Preservation Policy to the MMBW Maps Project, including image capture standards, image metadata and future proofing principles, we have ensured the long-term future of these newly-created digital resources.

This project is expected to be complete in early 2006, and has paved the way for many other heritage digitising projects at SLV. This includes the Port Phillip Papers Project, which involves the digitising of some of the Library's earliest and most important manuscripts relating to the European settlement of Victoria, and the audio-digitising of some of the Library's remarkable collection of 78 RPM sound recordings. The project has also contributed greatly towards the SLV's package of strategic initiatives, slv21, which aims to transform the Library's service model to one based around digital information and access appropriate to the 21st century. The adventure continues.