

Digital delivery: fast, convenient and long term

Tom Ruthven, The digitisation project of the National Library of Australia

Digital technologies and the internet offer libraries new opportunities to provide users with fast and convenient access to information and material, and the ability to reach new audiences. With this in mind, the National Library of Australia embarked last year on a major digitisation project to rapidly increase both the number of items available online and to broaden the range of material.

The project is digitising 20 000 pictures, 1000 maps and 3000 pieces of sheet music each year, and one manuscript collection at a time. This builds on the program of digitising pictures as they are acquired, which made 30 000 images available through the National Library's catalogue, Kinetica and PictureAustralia over the last six years.

Even with this increase there is a need to choose carefully what items to digitise. The library has a large, rich and diverse collection. Digitising 20 000 pictures will only touch the surface of the 600 000 that are held. Generally, the National Library digitises entire collections such as the Ferguson map collection or the Spencer collection of slides of Papua New Guinea. The key selection requirement is that the material is of significant research and cultural interest. Usually this means that it is original material or items which are unlikely to be held by any other library.

With a focus on providing as much material to users as possible, there needs to be minimal preservation work required and a preference for material that has already been catalogued or described. Finally, the image needs to be easy to create with the equipment that is available.

Curators choose collections or items based on their knowledge of its use and the research benefit to users. Orders for copies are analysed and users can suggest material or collections that they would like to see digitised through an online form. Each collection or item is assessed against the other criteria to see if, on balance, the digitisation of the collection will be relatively straightforward.

Copyright issues are also addressed in this process. Because the National Library is concentrating on unique items, the published material is usually older and falls outside any copyright restrictions. Sometimes, as with sheet music, there has been a deliberate decision to avoid copyright restrictions by digitising items that are out of copyright.

Permission to digitise and make available unpublished material is sought in a number of ways. With pictures, for example, the National Library has sought for many years the permission of owners to digitise and provide online access to the images as part of the acquisition process.

As well as providing access to the National Library's collection, digitisation enables us to collaborate with other libraries and cultural organisations to provide federated access to our collections. Users of the PictureAustralia service do not need to know the collection strengths or internet addresses of the eighteen contributing organisations. With one search they can find a picture from any contributor, breaking down the barriers between cultural organisations for users. This federated, collaborative approach is also being used with the MusicAustralia [pilot home page: <http://www.musicaustralia.org>] and AustraliaDancing [project home page: <http://www.nla.gov.au/australiadancing/>] projects.

While the benefits to users of direct access to library material from their workplace, school or home are relatively obvious, less apparent is the preservation role that surrogate copies have in managing and preserving a collection. A digital copy of a collection item

protects the original by absorbing the risk of damage that could be incurred through use. This is particularly important for paper-based materials published after 1850, many of which, printed as they were on acid-based paper, suffer the symptoms of brittle paper. Even moderate and careful handling can irreparably damage these items. It is also vital for rare or unique items where the consequences of damage could be irreparable. A copy of a digital item, made once and used many times, does not expose the original item to light or heat in the way copying of the original does.

This use of digital images as a surrogate has led some commentators to express concern at the durability of such digital objects, frequently citing the short life-span of CDs, data tapes and hard drives in support of their concerns. While it is indeed true that many of the modern information carriers do have limited useful life spans, it is equally important to note that it is not the carrier that libraries wish to preserve, but rather the information, or data, that it carries.

Digital information can be cloned without loss or degradation, whereas when the carrier of analogue information deteriorates, such as the unstable cellulose acetate or safety film used to produce many of the images of the 20th century, any copies to another analogue carrier introduces some loss of quality and, ultimately, information. However, once a digital image has been made, all future copies are an exact replica of that initial digital image, providing the potential to maintain the information content indefinitely.

The survival of a stream of data that is an exact and measurable replica of the original stream is well within the capabilities of existing technologies and systems. The real challenge is to ensure that the meaning of that data is not lost over time and to this end a responsible archive ensures that full description of all technical, structural and descriptive information is included in the form of explicit metadata. This will enable the routine management of systems for rendering the data, or provide clues for the digital archaeologist, should that ever become necessary.

Image files are amongst the least complex digital files to manage. A robust, uncompressed widely accepted image file, such as TIFF (Tagged Image File Format), presents relatively few problems in terms of long-term management. These files can be migrated to other equivalent formats without loss, and many software packages accept and read them. The imaging industry's investment in these files is large and any future changes in format will almost certainly be accompanied by software and systems that will enable bridging the differences, and if they should not, the task of producing appropriate software is a task that could be undertaken by many programmers. The real difficulty lies with the preservation of complex 'digitally born' files, such as are used to generate web pages and are stored on the National Library's PANDORA archive. These digital items must be archived as digital objects to maintain their integrity. The international library and archiving community is investing in considerable research and development to meet this challenge, and the solutions will help support the maintenance of digital image archives.

The critical concern with a digital archive is that it will not survive benign neglect; digital preservation is an active process. The contents of a digital archive will only be safe as long as there is a will, and the resources to support it. An investment in a digital archive is a long-term commitment.

Clearly the production of surrogate image files is an integral part of any library collection management strategy, and the benefits to users make digital duplicates the most appropriate option. As long as collection managers implement appropriate strategies these digital collections will remain functional and usable for the foreseeable

"It has revolutionised the way I find technical information.
When I have a question, it searches hundreds of IT titles
to quickly give me the correct, up-to-date answer."

SafariSM

TECH BOOKS ONLINESM

Search the best IT books online



Introducing SafariSM Tech Books OnlineSM, now available from ProQuest[®]. It is a fully-searchable database comprising the very best Information Technology content on the planet.

Hundreds of titles from the premier IT publishers are available on subjects ranging from XML to Networking to Java. It is cross-searchable, up-to-date and completely customisable.

IT answers in an instant

SafariSM
TECH BOOKS ONLINESM

For more information visit
www.proquest.co.uk
or contact

ProQuest
COMPANY

ProQuest Information and Learning, PO Box 181, Drummoyne, NSW, Australia 1470
AUST Free call: 1 800 11 6660 NZ Free Call: 0800 443429
email: sales@anz.proquest.com