Digitization in Australasia

This article surveys progress made by the libraries of Australia and New Zealand in planning and implementing major digitization projects. These projects exist within a wider context of the building and management of digital collections, including those developed through processes other than digitization.

New Zealand has developed a Digital Content Strategy which proposes a nationwide digitization programme based on key local, regional and national content. Australia is yet to develop a similar policy and funding framework.

In both countries, libraries face challenges in marshalling the resources and skills required to undertake large-scale digitization projects, and in addressing the issues which arise when they attempt to digitize in-copyright collection materials. The National Libraries have recognized the value in large-scale digitization of out-of-copyright newspapers, and have made significant progress towards establishing free, searchable newspaper archives.

WARWICK CATHRO
Assistant Director-General, Innovation
National Library of Australia

Development of digital collections

This article surveys progress with digitization in the libraries of Australia and New Zealand. It examines how digitization is being used to improve access to collections, and identifies some challenges which libraries face as they attempt to extend their digitization activities.

Digitization needs to be considered in a wider context, since it is only one of four strategies that libraries are using to build their digital collections. The other strategies are:

- acquisition of 'born digital' content through harvesting of web resources
- acquisition of 'born digital' content through deposit from authors or publishers
- licensing access to digital content supplied by vendors.

The harvesting of web resources has been undertaken in Australia through services such as the PANDORA archive[1] and the experimental capture of the entire Australian web domain[2]. Meanwhile the National Library of New Zealand is building the technical infrastructure for its planned National Digital Heritage Archive, which will collect born-digital publications under that Library's new legal deposit powers[3].

The deposit model of digital collecting is being used by Australasian universities as they build their institutional repositories through deposit of scholarly content such as articles, theses and data-sets. Two recent surveys of the state of development of institutional repositories in Australia have been conducted by the Council of Australian University Librarians (CAUL)[4,5].

Libraries have entered into agreements with vendors for access to digital content, often on a collaborative basis. CAUL has completed about 100 agreements with vendors for licensed access to content by Australian universities. New Zealand has established a national licensing arrangement across all library sectors known as EPIC (Electronic Purchasing In Collaboration) which is administered by the National Library of New Zealand[6]. The Australian library community has formed a National Licensing Forum, which developed a proposal for licensing access to a 'core set' of online content, such as major Australian academic and popular journals[7].

The other major strategy for digital collection building is digitization. Digitization has been defined as "the conversion of analogue information into digital information". In this context it refers to the conversion of library materials such as

books, pictures, maps, music scores, manuscript collections and audio and video files into digital form. It is useful to think of 'digitization' as encompassing all of the processes that are necessary to create usable digital files, including scanning or digital photography, metadata capture, creation of both master digital files and alternative versions (such as viewing copies and thumbnails) and the transfer of these files into a suitable storage environment.

The prime imperative for digitizing collection materials is to improve access to these collections by overcoming barriers of distance and time. Digitization enables people who live or work at significant distances from a collecting institution to view reproductions of the collection items through the web. Given faithful rendering of colour and use of advanced 'zoom' capabilities, the user's experience can match or even exceed the experience provided by access to the original item. When combined with effective discovery services, digitization provides virtually immediate access to the item.

Large-scale digitization of primary source materials can also provide a significant benefit to humanities and social science researchers by allowing them to explore, link and collaboratively annotate these research resources.

Digitization in Australia

In 1996 the National Library of Australia began to deliver access through its website to significant items in its pictures collection. In 2001 the Library embarked on a significantly expanded digitization programme which took in rare maps, early Australian sheet music and selected manuscript collections which form an important part of Australia's cultural heritage[8]. The Library has also been progressively digitizing its collection of more than 38,000 hours of sound recordings and has established the technical infrastructure required to deliver these over the web[9]. The Library has undertaken digitization in-house, and its digitization workflow is supported by a locally developed system known as Digital Collections Manager, which forms part of the Library's wider digital services architecture[10]. The Library's philosophy has been to provide an integrated approach, through its catalogue and other services, to the discovery of digitized collection items and to material not yet digitized.

Most of the Australian state libraries have digitized a selection of iconic items that record the history or culture of their state. For example, the State Library of Victoria has digitized over 200,000 pictures, historical manuscripts, maps and sheet music[11]. Funding of A$25M over four years was provided to the Library by the State Government in 2006 to support such initiatives and to extend the Library's online services. In July 2004 the State Library of New South Wales launched an initiative ('atmitchell.com') to digitize its iconic collections and make them available on the web[12].

The digitization activities of Australian university libraries have included development of e-reserve collections of course materials and digitization of examination papers[13]. Some university libraries hold heritage materials and are digitizing them for better access and preservation. For example, Deakin University has partnered with the National Library of Australia to digitize selected manuscripts and memorabilia of Alfred Deakin, Australia's second Prime Minister[14].

An early example of collaborative digitization in Australia involving universities was the Australian Co-operative Digitization Project. The material digitized was published originally in the period 1840-1845, a time which saw the emergence of an Australian colonial culture and identity, marked by an upsurge in local publication. Access to this digitized content continues to be provided by the National Library[15].

Another example of a major digitization programme undertaken in the Australian university sector is provided by SETIS, the Sydney Electronic Text and Image Service[16]. The service has published a number of digital collections, including an Australian Studies collection which has made available hundreds of out-of-copyright digital texts. The texts are re-keyed and given structural markup according to the standards for humanities source texts issued by the Text Encoding Initiative (TEI).

Australian cultural institutions other than libraries have also been active in digitization programmes. For example, the National Archives of Australia has digitized more than 18 million images and made them available through its website. In addition, the Archives has an active 'digitization on demand' service[17].

In 2006 the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) received special funding of A$12M to undertake a

major digitization project which aims to digitize up to a third of the Institute's audiovisual collections and to develop an 'online Indigenous research library'.

Digitization of local history and cultural collections can add significantly to the national resource. Some efforts have been made in Australia to provide guidance and assistance for such local digitization efforts[18,19].

Digitization in New Zealand

A survey of the digitization activities of over 300 New Zealand organizations was conducted in 2001[20]. The survey found that 39% of New Zealand libraries, museums and archives were currently digitizing elements of their collections. There was a strong emphasis on digitizing photographs, but other materials such as maps, plans and public records were also being digitized. The primary motive for digitization was to improve access to the collection materials.

The National Library of New Zealand has been prominent in the nation's digitization activities. For example, the 'Timeframes' service provides access to more than 42,000 images from the Alexander Turnbull Library, the division of the National Library of New Zealand which holds the nation's pre-eminent heritage collection. The Library has also digitized the Ranfurly Collection, which comprises paintings, drawings, papers and photographs compiled by the Governor of New Zealand from 1897 to 1904. In addition, the Library has undertaken a major newspaper digitization project, with an online delivery service called 'Papers Past', which is described in more detail below.

Digitization activity elsewhere in New Zealand was relatively modest at the time of the survey. Six academic libraries, eight public libraries and 12 museums were undertaking projects in 2001.

Digitization of newspapers

The digitization of extensive runs of out-of-copyright newspapers has the potential to create a rich historical resource for social science researchers, family historians and the general public.

The National Library of New Zealand has developed Papers Past[21], a collection of digitized 19th-century newspapers and periodicals comprising close to one million pages, freely available online. The Library commissioned Preservation Resources, a US-based, not-for-profit organization, to undertake the scanning, and it later commissioned Planman Technologies to undertake Optical Character Recognition (OCR) and other conversion services which will allow Papers Past to support full text searching of the newspaper database.

The National Library of Australia is planning a project to digitize nine major Australian newspapers, covering out-of-copyright issues from 1803 to 1954, which comprise more than three million pages. These pages will be made fully searchable and browsable through the use of OCR and other conversion services. The Library initially partnered with the Australian National University and 12 other universities to prepare a funding bid to the Australian Research Council. The bid received strong support from the state libraries, the Australian Academy for the Humanities and individual academics, but was unsuccessful. Nevertheless, the Library has been able to proceed with the project using other funds in its budget, and was expected in early 2007 to announce a contract for the conversion services needed to make the newspaper database a reality.

Service frameworks

Digital collections must be managed over time and must be capable of being effectively discovered and delivered to users. At the institutional level, it is desirable to manage the digital collections in conformity with the Open Archival Information System model[22] and to uncouple the processes of managing and discovering the resources from the process of acquiring them in digital form. This means that processes such as rights management can be applied to all digital objects however they were created or ingested.

In Australia, there has been a strong awareness of the need to identify digital collection items through 'persistent identifier' systems and resolution services that allow links to these collection items to function reliably notwithstanding changes to the location of the resources which those links cite. The National Library of Australia has developed its own persistent identifier scheme and resolver service[23] and has

issued guidelines on the management of web resources for persistent access[24]. In addition, a project known as PILIN (Persistent Identifier Linking Infrastructure) was funded in 2006 under the Australian Government's Systemic Infrastructure Initiative[25].

Digital collection items should be capable of being discovered alongside other collection items that are not yet digitized. For this reason they are normally included in library catalogues and other discovery services such as the union catalogues provided by the National Libraries.

Users will prefer not to repeat their search for digital collection items on multiple websites. In response to this requirement, federated discovery services have been developed. In addition to the national union catalogues, these services include PictureAustralia and MusicAustralia, services hosted by the National Library of Australia[26,27]. To further the goal of making discovery easier, the Library is taking steps to expose the metadata used by these services to Internet search engines such as Google[28].

Typical browse requirements are the need to move up and down a collection hierarchy, to move sequentially through the pages of a multi-part item (such as a digitized diary or music score) and to zoom in to examine detail (such as a part of a digitized map). The National Library of Australia has implemented these requirements as part of its digital collection architecture[10] and most of these functions are also available through commercial products such as DigiTool from Ex Libris[29].

Policy and funding frameworks

Collection digitization is expensive. The funding to support digitization may come from a variety of sources such as a library's existing budget, a government grant (including grants made through research infrastructure programmes), a commercial partnership, or a foundation or other sources of private philanthropy. Digitization of out-of-copyright works through public sources of funding can enable such works to be widely and freely available in their online form, unrestricted by licences which apply when they are digitized by the commercial sector.

Ideally, there should be a government policy framework that recognizes the community benefit in digitization and provides some recognized

avenues for funding at least some projects which meet defined criteria for delivering that benefit. Such a framework does not yet exist in Australia, though there has been more progress in this area in New Zealand.

The New Zealand Digital Strategy[30] was launched in May 2005 following an extensive public consultation process and the involvement of more than 50 government agencies. The Strategy aims to maximize the benefit to New Zealanders from the development of digital content and infrastructure. It includes a Digital Content Strategy which, among other things:

- recognizes that digitizing New Zealand's existing collections will be an important means of making that content more visible and of enabling broader public access
- proposes a nationwide digitization programme based on key local, regional and national content
- proposes the provision of support and advice to communities on the standards and tools needed to create and share content.

There is nothing in Australia that approaches this policy statement of the New Zealand Government. The efforts to achieve a similar national vision in Australia have been uncertain and fragmented.

In 2003, a committee of the Australian Senate issued a report[31] on *Libraries in the online environment*. Among other things, the report proposed a 'new National Heritage Grants programme for peak cultural institutions to assist in the digitization of their collections'. However, no action has been taken to implement this recommendation. Furthermore, collection digitization has been ruled out of scope for the various programmes that exist to fund research infrastructure.

Some of those seeking a national digitization fund are now looking to the outcome of the Digital Collections Summit, held in August 2006[32]. That event was convened by the Collections Council of Australia, the peak body for the Australian collections sector, covering archives, galleries, libraries, museums and other collecting institutions. The Summit brought together a wide range of representatives of those sectors, with the aim of presenting a programme of action to the Cultural Ministers Council, which comprises ministers of the Australian Government and state, territory and New Zealand governments with responsibility for the arts. The need for a national digitization fund

was one of the issues identified during the Summit. The final report of the Digital Collections Summit is expected in early 2007.

Copyright issues

Digitization projects are commonly scoped with a view to avoiding potential infringement of copyright, or avoiding the complexities that arise from uncertainties in this area. For example, libraries may seek to avoid copyright issues with future digitization projects by negotiating permissions at the time of acquisition.

Libraries that undertake such digitization projects may seek to obtain a licence permitting this. For example, the National Library of Australia, after digitizing Australian music scores in its collection that were published before the mid 1930s, has endeavoured to find a legally sound approach to digitizing those scores published more recently. As the task of identifying and contacting the current copyright owners would be practically impossible, the Library has sought to pay for a licence to undertake this project. Persistent approaches to the music publishers and to the Australasian Mechanical Copyright Owners' Society (AMCOS) have failed to identify any efficient process to provide the Library with the relevant licence. Such a licence would benefit not only the users (music researchers, students and the general public) but also the copyright owners, who have not made any revenue from these publications in several decades.

In some circumstances, libraries and archives may be able to proceed with a digitization project after making a suitable form of public advertisement. One of the collections digitized by the National Library of Australia was the Barton Papers, the archive of personal papers of Australia's first Prime Minister. Rather than attempting to obtain permission from each of the heirs of the more than 400 correspondents in the collection, the Library was able to proceed under the terms of Section 52 of the Copyright Act 1968, which deals with the publication of unpublished works kept in libraries or archives. The Library placed a notice of intention to publish in the *Government gazette*, citing all 400 names, and proceeded with the project after no objections were received[33].

Another potentially complex issue concerns the relative rights of publishers and authors. This issue can arise in cases where a library wishes to digitize back-runs of scholarly journals that are still in copyright. The publisher of the journal may take the view that it is unable to grant a licence to digitize because there has been no explicit transfer of this right to the publisher. Given that the authors of articles written several decades ago may be difficult to contact, libraries and publishers may agree to proceed with digitization subject to a public disclaimer under which the library would withdraw online access to any item on request by the author, or else negotiate conditions under which access would be retained.

Conclusions

The libraries of Australia and New Zealand have been active in recent years in planning and implementing major digitization projects. These projects sit within a wider context of the building and management of digital collections, including those developed through processes other than digitization.

New Zealand has been successful in developing a Digital Content Strategy which proposes a nationwide digitization programme based on key local, regional and national content. Australia is yet to develop a similar policy and funding framework.

In both countries, libraries face challenges in marshalling the resources and skills required to undertake large-scale digitization projects and in addressing the issues which arise when they attempt to digitize in-copyright collection materials.

The National Libraries in both countries have recognized the value in large-scale digitization of out-of-copyright newspapers, and have made significant progress towards establishing free, searchable newspaper archives.

References

1. PANDORA: Australia's web archive:
   http://pandora.nla.gov.au/index.html

2. First whole Australian web domain harvest, *National Library of Australia Gateways*, 78 (December 2005):
   http://www.nla.gov.au/pub/gateways/issues/78/story01.html

3. National Digital Heritage Archive Programme –
    Q&As:
    http://www.natlib.govt.nz/en/whatsnew/
    4initiatives.html

4. Shipp, J, *Making the case for institutional repositories:
    Country update – Australia*:
    http://www.caul.edu.au/surveys/InstRepositories
    2005.doc

5. Richardson, J, *Integration of open access repositories
    with research management systems [CAUL survey]*.
    August 2006.
    http://www.caul.edu.au/surveys/OARintegration
    2006.doc

6. EPIC – Electronic Purchasing In Collaboration:
    http://epic.org.nz/nl/epic.html

7. National Licensing Proposal:
    http://www.nla.gov.au/nlp/

8. Digitization – overview:
    http://www.nla.gov.au/digital/

9. Ayres, M-L, Pearce, J and Dahlitz, D, *Bringing the
    stories to the people: online sound at the National Library
    of Australia*:
    http://www.nla.gov.au/nla/staffpaper/2006/
    documents/Mari-Louise.pdf

10. Cathro, W and Boston, T, *Development of a digital
    services architecture at the National Library of Australia:*
    http://www.nla.gov.au/nla/staffpaper/2003/
    cathro1.html

11. State Library of Victoria. Digitizing the collection:
    http://www.slv.vic.gov.au/about/pastfuture/future
    /digitizing/index.html

12. Thundatil, N, *Architecture for implementing a
    sustainable electronic delivery services platform –
    atmitchell.com case study*:
    http://conferences.alia.org.au/alia2006/Papers/
    Nandu_Thundatil.pdf

13. Horn, A and Schmidt, J, *Digitization activities of
    Australian university libraries*:
    http://www.nla.gov.au/initiatives/alctf/meeting9/
    agenda_item12.html

14. Signing over Deakin: collaborative project for
    Deakin Papers website. *National Library of Australia
    Gateways*, 63 (June 2003).
    http://www.nla.gov.au/pub/gateways/archive/
    63/p15a01.html

15. Australian Co-operative Digitization Project,
    1840–1845:
    http://www.nla.gov.au/acdp/

16. The Sydney Electronic Text and Image Service
    (SETIS):
    http://setis.library.usyd.edu.au/

17. Ling, T, *Taking it to the streets: why the National
    Archives of Australia embraced digitization on demand*:
    http://www.naa.gov.au/publications/corporate_
    publications/digitizing_tling.pdf

18. Image digitization of local history collections:
    Victorian public libraries manual.
    http://www.libraries.vic.gov.au/downloads/
    Victorias_Virtual_Library_Digital_Collection/
    lochist.htm

19. Collections Australia Network. *Capture your
    collections: a digitization course for regional museums
    and galleries*:
    http://www.collectionsaustralia.net/sector_info_
    item/10

20. Dorner, D G, Chawner, B and Searle, S, *Creating
    a Digital Babylon? Results of a survey of digitization
    activities in New Zealand*. A report commissioned
    by the National Library of New Zealand:
    http://www.natlib.govt.nz/files/4initiatives_
    digsurvey.pdf

21. Papers Past: New Zealand newspapers and
    periodicals:
    http://paperspast.natlib.govt.nz/

22. Reference model for an Open Archival Information
    System (OAIS):
    http://public.ccsds.org/publications/archive/
    650x0b1.pdf

23. Boston, T and Nguyen, N, *A practical approach to
    ensuring the persistence of digital collections at the
    National Library of Australia*, 2002.
    http://www.nla.gov.au/nla/staffpaper/2002/
    boston2.html

24. National Library of Australia. *Managing web
    resources for persistent access*:
    http://www.nla.gov.au/guidelines/persistence.
    html

25. PILIN Project:
    http://www.arrow.edu.au/PILIN.php

26. Hooton, F, *The big vision – Picture Australia*:
    http://www.nla.gov.au/nla/staffpaper/2006/
    documents/Big-Vision.pdf

27. Holmes, R and Ayres, M-L, MusicAustralia*: towards
    a national music information infrastructure*:
    http://www.nla.gov.au/nla/staffpaper/2004/
    ayres2.html

28. Boston, T, *Exposing the deep web to increase access to library collections*:
http://www.nla.gov.au/nla/staffpaper/2005/boston2.html

29. DigiTool:
http://www.exlibrisgroup.com/digitool.htm

30. The Digital Strategy: creating our digital future:
http://www.digitalstrategy.govt.nz

31. Australia. Senate. Environment, Communications, Information Technology and Arts References Committee. *Libraries in the online environment.* October 2003:
http://www.aph.gov.au/Senate/committee/ecita_ctte/completed_inquiries/2002-04/online_libraries/report/report.pdf

32. Collections Council of Australia. Digital Collections Summit:
http://www.collectionscouncil.com.au/summit+2006+-+digital+collections.aspx

33. Waring, P, *Edmund Barton digitization project*:
http://www.nla.gov.au/nla/staffpaper/2001/waring1.html

■ **Warwick Cathro**
Assistant Director-General, Innovation
National Library of Australia
Parkes Place
Canberra, ACT 2600, Australia
Tel: +61 (0)2 6262 1403
Fax: +61 (0)2 6273 1133
E-mail: wcathro@nla.gov.au

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=1&spage=9

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=20&issue=1