

2011 Strategic Roadmap for Australian Research Infrastructure Discussion Paper

Department of Innovation, Industry, Science and Research

Submission:

National & State Libraries Australasia

www.nsla.org.au

nsla@slv.vic.gov.au



This submission addresses issues outlined in *Chapter 5: Understanding Cultures and Communities Expert Working Group*. Responses are provided to the following:

- 5.A.2 How should we **prioritise research areas** for this sector when developing Australia's research infrastructure?
- 5.B.1 What are your views on the **research infrastructure capability areas** identified, including their relative priority and their ability to support the current and future research needs?
- 5.F.1 Are there other programs/issues/developments not listed that you consider could be a driver for **future research infrastructure investments** or may impact on such investments?

Part 1. Summary and Recommendation

All information is now expected to be online. NSLA represents the National, State and Territory Libraries of Australia and New Zealand and the combined collections held in these ten major publicly-funded research libraries represent a significant portion of our nations' research infrastructure. These collections are our irreplaceable assets, gathered and invested in by governments for more than 150 years and currently valued at more than AUD5 billion. To deliver a return on this investment, these collections must be digital and available for access via the internet.

To the maximum extent practicable, information research and content funded by Australian governments – including national collections – should be made freely available over the internet as part of the global public commons

Venturous Australia, Review of the National Innovation System
Dept of Innovation, Industry, Science & Research, 2008

An eResearch infrastructure Capability supporting the humanities, arts and social sciences was outlined in the 2008 Roadmap but was not ultimately funded. The Capability featured two broad and connected elements of eResearch infrastructure:

- (1) *data creation – through digitisation, systematic capture of 'born digital' materials, and support for national survey instruments; and*
- (2) *data management – including curation and dissemination through coordinated strategies and platforms.*

2011 Roadmap for Australian Research Infrastructure, Discussion Paper, p.46
Dept of Innovation, Industry, Science & Research

Digitise the collections of Australia's major national institutions by 2020
"Top Ideas – Creative Australia Stream", **Australia 2020 Summit**,
Government Response, April 2009.

These three statements represent the core research infrastructure recommendations of National & State Libraries Australasia (NSLA). Libraries are deeply engaged in the shift of information to digital form, and the preservation and access challenges of digitised and born-digital materials.

The Australian Research Council's *Excellence for Research in Australia 2010 National Report* (See Discussion Paper, p. 8) highlights the areas of research where Australia performed "well above world standard". Within this is the particularly broad field of "historical studies" which underpins research in the humanities and many areas of the sciences. NSLA Libraries hold Australia's documentary heritage, the primary research content for Australian historical studies.

The availability of the historical record, increasingly in digitised form, is one element of the scientific dominance the United Kingdom holds in the scientific discourse on climate change
Inspiring Research, Inspiring Scholarship,
Joint Information Systems Committee (JISC), UK, October 2010

The scale of digitisation required in Australian libraries is enormous, though the funding required to make important progress in the next five years is comparatively small. NSLA advises that AUD100 million will shift Australia to a mass-digitisation model and begin to make our significant assets visible and available online. Within this, there are groups of information assets, in particular newspapers, that have broad application for research and can be separated for immediate digitisation to create high-use datasets.

Australian Newspapers, previously available only on microform or in decaying bound volumes, began to become available online in August 2008 through Trove (trove.nla.gov.au), the national portal to Australian content. Since then *Australian Newspapers* has become the most accessed part of the TROVE website. As at April 2011, there are almost 50 million fully-searchable articles from selected newspapers available online.. Currently the content is primarily from one major metropolitan newspaper from each state and territory across Australia. Another 10 million pages from a wider range of newspapers are ready to be digitised when funding is available.

Recommendation

As a first step, NSLA recommends investment in increased digitisation of Australia's newspapers to expand and enhance Trove as a rich core dataset for multi- and cross-disciplinary research.

NSLA also supports the recommendations of the National Library of Australia, the Council of Australian University Librarians (CAUL) and Australian Library and Information Association (ALIA) to develop Australia's research capability, and continuing investment in the Australian National Data Service (ANDS).

Part 2: Prioritising Research Areas

The use of *Australian Newspapers* online has outstripped all expectations – with use and re-use across an extensive range of research fields, by academic, government and independent researchers.

In the first 3 years of the service (to April 2011), there were more than 3 million unique users of the online newspapers site – a far greater number than could ever have accessed historic newspapers in their physical format. Each day, approximately 20,000 people use the service. Of these, more than half of the users identify themselves broadly as 'researchers'. As materials become available online, research opportunities progressively open up. Digitised resources transform the research process; new areas of research are enabled; new questions can be asked that were previously not feasible.

Rather than prioritising the significance and complexity of research in humanities subject areas, NSLA supports growing a core dataset that can be mined by researchers across the humanities and sciences.

Some case studies that highlight both the wide use of newspapers via Trove and also the crowd-sourced text correction that continues to improve the online search functionality:

1. **South East Australian Climate History [SEARCH], ARC Linkage Project LP0990151 2010-2012**, led by University of Melbourne (Professor David Karoly and Dr Joelle Gergis) with partners including the Australian Bureau of Meteorology, Murray-Darling Basin Authority, Department of Sustainability and Environment (Victoria), National Library of Australia, State Library of Victoria, and others. This project has used online newspapers extensively, engaging a volunteer workforce through libraries to trawl for standard metrological data and for the record of extreme weather events.

“Australia’s early settlers also battled plagues of insects and extreme weather, according to accounts from the country’s first newspaper, *The Sydney Gazette and New South Wales Advertiser*. A team of volunteers at the State Library of New South Wales has been hard at work extracting weather-related articles from the newspaper, along with meteorological readings from the early years of the colony at New South Wales. These documentary accounts are being used by the SEARCH research team to help extend South Eastern Australia’s climate record back to 1788...proving to be hugely beneficial to SEARCH’s research.” www.climatehistory.com.au
2. A doctoral researcher at the University of Sydney has used Trove extensively to precisely date the introduction of pieces of **sheet music** into Australia between 1825 and 1855, using articles and advertisements published in the historical newspapers (now digitised). The results will be available via his thesis as well as various journal articles, and he now intends to extend this to the turn of the nineteenth century.
3. A lecturer at the University of Canberra has been developing *An analysis of critical factors that influence the outcomes of Gov 2.0 initiatives within Australian government from a collaboration perspective*, including an investigation of crowd-sourcing for text-correction used by Trove for newspapers online.
4. **Oxford University Press** uses online newspapers to determine when certain words were first used in Australia: www.oup.com.au/dictionaries/wotm/neenish_tart.

Part 3: Research infrastructure capability areas

Digitisation

To date, digitisation has concentrated on selected parts of libraries’ heritage collections of pictures, manuscripts, maps and historic newspapers, with great achievement, though still only a fraction has been made available online. Australia lags many other countries in directing extensive public funding to mass digitisation and the pressure to address this is growing. Failing to digitise will result in a much-reduced Australian research and cultural presence on the web.

There is recognition in Australia and in many other countries of the importance of digitisation in improving educational, social and employment opportunities. Many countries have committed significant funding:

- France committed more than AUD1 billion in December 2009 for the digitisation of French-language books;
- The Japanese government has provided a supplementary budget of AUD125 million to digitise 900,000 items in 2009-10;
- The German Digital Library (DDB) will come online in 2011;
- Norway has announced in 2009 that they will digitise the entire collection of the National Library of Norway;
- In 2010, the National Library of the Netherlands announced its plans to digitise all Dutch books, newspapers and periodicals from 1470 onward;

- In May 2010, the British Library announced it will be working with publisher Brightsolid to digitise 40 million newspaper pages from 52,000 newspapers from across the United Kingdom, this is one of several UK digitisation initiatives;
- On 1 June 2010, the New Zealand government committed AUD9.1 million to develop a Government Digital Archive; and
- Government and philanthropic donations in the United States are supporting digitisation of the Library of Congress and other collections, including the American Memory, National Digital Newspaper Program and other online collections (for example, a grant from the Sloan Foundation to digitize books in the public domain has resulted in the digitisation of more than 52,600 volumes, totaling approximately 9.3 million images by September 2009).

The Librarian of Congress, James H. Billington, interviewed in the [New York Times](#) in January 2011, announced that planning has commenced on a US national digital library, “a gigantic digital library that would make the cultural heritage of the country available to everyone”. He notes that the US is “playing catch-up ... as the United States finds itself trailing Europe and Japan ... [creating] a real digital library divide”.

Built up over centuries at an enormous expenditure of money and labor, library collections can be digitised en masse at relatively low cost – millions certainly, but little compared to the investment that went into them.
Professor Robert Darnton, Harvard University, 2010

Digital preservation

...economically sustainably preservation - ensuring the ongoing and efficient allocation of resources to digital preservation – is an urgent societal problem. It is urgent because digital information is inherently fragile, prone to information loss and degradation. Preservation insures against multiple risks to information assets over time.
Blue Ribbon Taskforce on Sustainable Digital Preservation Access, USA, 2010

Along with the digitisation of our collections, a central and inter-linked challenge is the long-term preservation of digital assets whether born-digital or digitised from another format. The processes to digitise are tested and robust and all that is required to gear up to mass digitisation is funding. Preservation of digital assets is the more complex challenge. To both preserve and provide access in the future to digital assets created on many generations of technology platforms is an evolving and approaches.

The digital content that requires preservation for the future is growing exponentially, encompassing more types of materials and collections. It includes personal, research and company archives that are now predominantly born-digital, as well as the need to select and preserve web content. Websites are fluid with constantly changing content and only a small fraction are currently selected at designated intervals and preserved for easy access. Much valuable content is lost as the content churns.

Investigation and testing of models and standards for preservation is underway in many countries and Australia must move to ensure its digital assets are maintained in the interim and that appropriate solutions for content are adopted before critical digital content is lost.

Part 4: Future research infrastructure investments

Investment in the digitisation of, and online access to, newspapers is recommended as a first step in building improved research infrastructure. The systems of work for creating this digital content are established; processes are tested and mature; and the resulting use has proven to be extensive across the Australian research community.

NSLA Libraries have 10 million additional newspaper pages selected and ready for digitisation when funding is available, at a cost of approximately AUD2.96 per page. The process can be managed by the National Library of Australia with a short lead time and quickly expand and enhance this high-use digital asset.

Appendix 1

Members of National & State Libraries Australasia:

ACT Library & Information Service	State Library of Queensland
National Library of Australia	State Library of South Australia
National Library of New Zealand	State Library of Tasmania
Northern Territory Library	State Library of Victoria
State Library of New South Wales	State Library of Western Australia

In 2009-2010:

11.4 million	people visited our libraries;
233.4 million	pages were viewed on our websites;
1,018 terabytes	of digital collections were stored;
\$36 million	was spent purchasing resources for collections;
\$5.3 billion	was the total asset value of our collections; and
\$1.06 billion	was the total asset value of our buildings/sites.



Mr Alan Smith

Director, State Library of South Australia &

Chairman, National & State Libraries Australasia

www.nsla.org.au | nsla@slv.vic.gov.au

May 2011