

# **Beyond Discovery: Digital Scholarship, Connected Communities and the Evolution of Trove**

**Dr Tim Sherratt**

National Library of Australia, Canberra, Australia, [tsherratt@nla.gov.au](mailto:tsherratt@nla.gov.au)

For centuries libraries have supported research by cataloguing, compiling and collecting – by helping researchers find and use material relevant to their studies. But as these compilations become available in digital form they provide more than aids to discovery, they themselves become the raw material for digital analysis and transformation – they become sites for the creation of new knowledge.

Trove, the National Library of Australia's discovery service, provides access to millions of collection records drawn from libraries, museums, archives, government agencies, universities, and other research organisations. It also includes a rich collection of fully-digitised Australian newspaper articles. Increasingly this data is being used by researchers in new ways through the application of computational methods and the development of innovative digital tools. This paper will explore this trend and consider its implications for discovery services such as Trove.

## **WHAT IS TROVE?**

Trove is an open national collection which documents Australian life and society. Built on three decades of technological collaboration, Trove helps researchers engage with rich digital content, share their knowledge for the benefit of others, and re-use public domain content to meet both research objectives and community needs. The National Library brought the Trove data corpus together with a simple, single discovery premise – a concept that underpinned services such as the Australian Bibliographic Network, Picture Australia, Australian Research Online and the People Australia project, all now part of Trove.

The Trove corpus is compiled from several significant data sources including the National Bibliographic Database managed by Libraries Australia, more than 90 million digitised newspaper articles from historical newspapers 1803-1954, article metadata for more than 100 million articles provided by Gale and RMIT Publishing, collection information from the GLAM (galleries, libraries, archives and museums) sector, and all institutional repositories in the Australian higher education sector. The repositories include digital theses and open access journals which Trove pools into a single informative source for Australian research.

## **NEW RESEARCH USES**

Around the world bibliographic data, the digitised contents of books and journals, and the holdings of cultural institutions are being subjected to new forms of digital analysis. Research impact and networks of influence can be explored through citation analysis. The mining of large text collections can reveal intellectual and cultural trends. Visualisations of cultural collections can offer new modes of engagement and discovery. Access to large quantities of digital data, supported by the development of analytical tools and computational techniques, allows new research questions to be framed.

Trove's collection of digitised newspapers already supports a surprising range of research projects. The ARC-funded Australasian Colonial Legal History Library, for example, has used the database to recover case reports not available elsewhere. [1] Climate researchers have been mining the newspapers for information on weather patterns, while the University of Tasmania has used them to find clues about the arrival of invasive insect species. A major ARC-funded project at the Australian National University is exploring the characteristics of serialised fiction in Australian newspapers using large-scale data harvesting and enrichment.

While the digitized newspapers have been most heavily used, other collections offer interesting possibilities for large-scale analysis. Following Ben Schmidt's experiments in graphing trends across American history theses, it is

easy to imagine how Trove's collection of digital theses might be interrogated to explore patterns in Australian intellectual history. [2]

To support digital scholarship, data providers, like Trove, have to consider the most effective ways of delivering data subject to a range of technical and legal constraints.

## **INNOVATIVE TOOLS**

Trove provides researchers with familiar search and browse interfaces as well as a number of annotation options. However, it obviously cannot predict or support all of the ways researchers might want to interact with its data. Like similar services including Europeana, the Digital Public Library of America and DigitalNZ, Trove provides an API to enable researchers and developers to access its data through their own tools and applications. More than just a portal for discovery, Trove offers a platform for development.

QueryPic, for example, is a freely-available web application that enables users to visualise searches in the Trove newspaper database over time. [3] Developed using the Trove API, QueryPic makes simple digital research techniques available to a wide audience. Other projects, such as Mitchell Whitelaw's TroveMosaic, use the API to investigate approaches to visualisation and the possibilities of 'generous interfaces'. [4] The Trove API was also heavily used during GovHack 2013, resulting in a number of innovative applications such as Pixtory and Advintage that explored ways of engaging the public with cultural heritage collections.

Publishing and maintaining an API brings data providers into contact with a new audience – developers – whose needs and concerns might be quite different to those of traditional users. How does a service like Trove engage and support them?

## **ENGAGED COMMUNITIES**

A passionate and committed community of users has developed around Trove, most evident in the efforts of volunteers who correct the OCR output of articles in the newspaper database. Annotation tools, such as tags and lists, also enable users to build their own collaborative research projects.

But the Trove community itself is increasingly of interest to researchers interested in online behaviour. Sultana Lubna Alam, for example, has used Trove in her examination of crowdsourcing motivations in the GLAM sector. [5] Annie Talve's collective intelligence and participatory design research has similarly explored the value provided by passionate people to large scale online aggregation services. [6]

Through sharing, re-use, citation, and annotation, the content of Trove will increasingly be discovered and consumed beyond the National Library of Australia's own website. The online conversations and communities that will form around the content have the capacity to greatly enrich the raw metadata if they can find and harvested. Already a WordPress plugin exists that makes it easy to automatically embed newspaper citations using the Trove API. As scholarly publication itself evolves to take advantage of the online environment, services such as Trove will be challenged to become not just data providers, but active participants in these connected communities.

## **REFERENCES**

1. Greenleaf, G. et al, 'Digitising and searching Australasian colonial legal history', Presented at Australian Historical Association 31st Annual Conference, Adelaide, 2012 <<http://www.theaha.org.au/conferences.html>>.
2. Schmidt, B. , 'What years do historians write about?', <<http://sappingattention.blogspot.com.au/2013/05/what-years-do-historians-write-about.html>>.
3. QueryPic, <<http://dhistory.org/querypic/>>.
4. TroveMosaic, <<http://mtchl.net/TroveMosaic/>>.
5. Alam, S. and Campbell, J., 'Crowdsourcing motivations in a not-for-profit GLAM context: the Australian newspapers digitisation program', Proceedings of the 23rd Australasian Conference on Information Systems, 2012, <<http://dro.deakin.edu.au/eserv/DU:30049107/alam-crowdsourcingmotivations-2012.pdf>>.
6. Talvé, A., 'Do YOU Trove?', <<http://www.projectsusu.com/2013/03/do-you-trove/>>.

**TIM SHERRATT**

Dr Tim Sherratt is the Manager of Trove at the National Library of Australia. He is a digital historian and cultural data specialist who has been developing online resources relating to history, archives and museums since the mid-1990s. He is also an Adjunct Associate Professor in the Centre for Creative and Cultural Research at the University of Canberra.