

[HOME](#) [BLOG](#) [ABOUT](#) [UPDATES](#) [PROJECTS](#)


Tim Sherratt
Historian and hacker

Trove: connecting us to the past

by Tim Sherratt
6 July 2019

CATEGORIES [Trove](#)

Published in *Teaching History* (Journal of the History Teachers Association of NSW), vol. 53, no. 2 (June 2019).

Ask people to describe [Trove](#) in a single word, and they we generally answer 'newspapers'. And why not? By making more than 200 million Australian newspaper articles from 1803 onwards available online, Trove has changed the practice of history in ways we don't yet understand.

But as revolutionary as Trove's newspapers have been, there's much more to explore. Created by the National Library of Australia, Trove is a digital repository, a collection of collections, and a platform for collaboration and creativity. Not only is it an important starting point for any historical research, it raises critical questions about how digital resources are changing our access to the past.

Not just newspapers

Before we move on from the newspapers, it's worth reflecting on sheer variety of titles that have been digitised and made easily accessible online. You won't just find the familiar metropolitan dailies, there's hundreds of small regional newspapers as well. There's political papers, religious papers, community papers in French, German, Croatian, and even Chinese. One of my favourites is the *Chinese Advertiser* (later the *English and Chinese Advertiser*), which was published on the Ballarat gold fields in the 1850s. Only ten issues are known to survive and they're all in Trove.



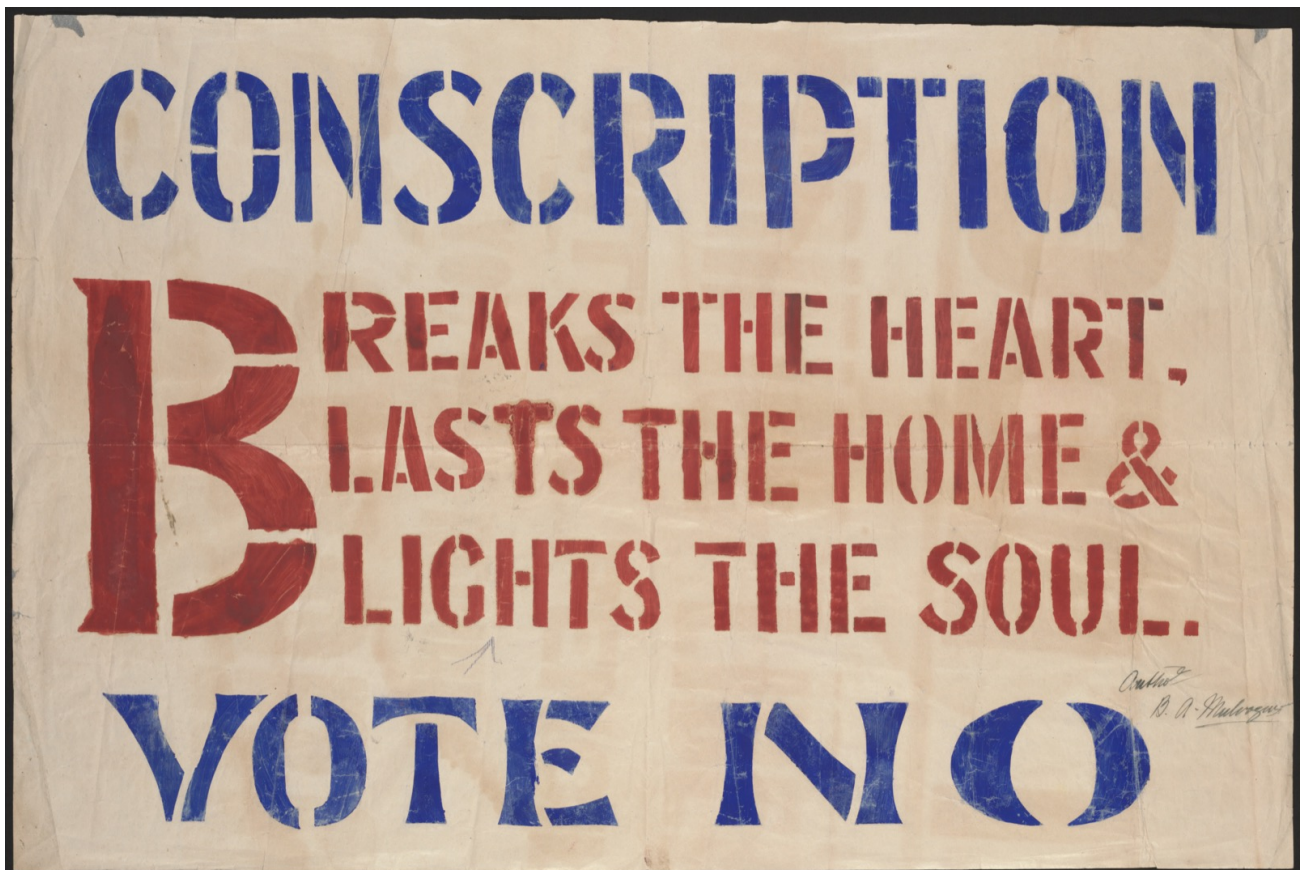
<http://nla.gov.au/nla.news-page15910494>

The digitisation process creates images of each newspaper page. Then, through a process known as Optical Character Recognition (OCR), it breaks the page up into separate articles, and extracts their textual content. By turning images into text, Trove makes those 200 million articles searchable.

But outside of Trove's newspaper zone, there are many other digitised resources to explore online. In the journals zone, for instance, you'll find more than 4,000 issues of *The Bulletin* from 1880 to 1968. In the late nineteenth century, *The Bulletin*'s mix of politics and literature fostered both a new sense of radical nationalism, and support for the racist ideal of 'White Australia'. Other digitised journals include the NSW *School Magazine of Literature for our Boys and Girls* from 1919 to 1954, the *Pacific Islands Monthly* from 1930 to 2000, the *Australian Woman's Mirror* from 1924 to 1954, the *Wireless Weekly* from 1922 to 1943, and even the *Kennel Control Council's gazette* from 1932 to 1969!

There are more than 300 digitised journals in Trove which, like the newspapers, can be searched at the level of individual articles. However, it's currently quite difficult to search within a particular journal on Trove. To help with this, I created a simple web app that helps you [to explore the digitised journals](#) – just select the titles you're interested in, insert some keywords, and the app will build your search and redirect you back to the results in Trove.

Other Trove zones – books, pictures, and maps – also contain large volumes of digitised material that you can view and use online. Amongst the books you'll find a fascinating collection of ephemera – things like leaflets, posters, pamphlets, certificates, and even playbills. There are, for example, a number of items relating to wartime recruiting and the WWI conscription debate.



<https://trove.nla.gov.au/nla.obj-508114389/>

The map zone is particularly rich with freely available, digitised content. I did a bit of [poking around recently](#) and worked out that there are more than 20,000 maps available online, and about 80% of these are free of copyright restrictions. Like other digitised resources, the maps cannot only be viewed online, you can download high-resolution copies for printing or sharing. In the case of textual resources, like books and journals, you can download complete issues, individual page images, or the raw text extracted by OCR processing – just look for the download tab in the digital resource viewer.

However, as with the journals, finding all of this exciting digitised content can be difficult. There's just so much stuff in Trove! One simple trick is to include the term "nla.obj" in your search. Digitised resources are given a unique identifier that includes this string, so you can use it to filter your results. If you're interested in getting deep into the collection data, have a look at my [GLAM Workbench](#). There you'll find sections on Trove's books, journals, and maps that include lists of resources and well as bulk collections of full text. Yes, that includes [6.6gb of text extracted from 27,426 journal issues!](#)

We tend to use terms like 'digital' and 'digitised' interchangeably, but a large proportion of Trove's digital resources didn't go through any sort of digitisation process – they were just born that way! 'Born digital' works include a growing number of recent journals and books that were published electronically and are available online in formats like PDF or MOBI. But by far the biggest collection of born digital resources can be found in Trove's archived websites zone. There'll you'll find an astonishing 5 billion snapshots of Australian websites since 1996.

But why would historians be interested in a collection of websites? Hmmm, perhaps we should reverse that question. How will future historians understand the 1990s or beyond without using web archives? Collections like these capture a wide range of social, political, and cultural activity that won't have been documented anywhere else. As we know, websites change and disappear all the time, whether by design or accident. Web archives capture these changes, enabling us to look back in time, treating the internet itself as a historical source.

Trove's web archive raises another important question, just how do you make sense of collections on this sort of scale? What does it mean when you search for something in Trove's newspapers and discover there are two million matching results? Digitisation has unlocked resources, enabling new forms of discovery and use, but with improved access has come the challenge of abundance. Just what does it all mean?

The Great War vs First World War & World War I



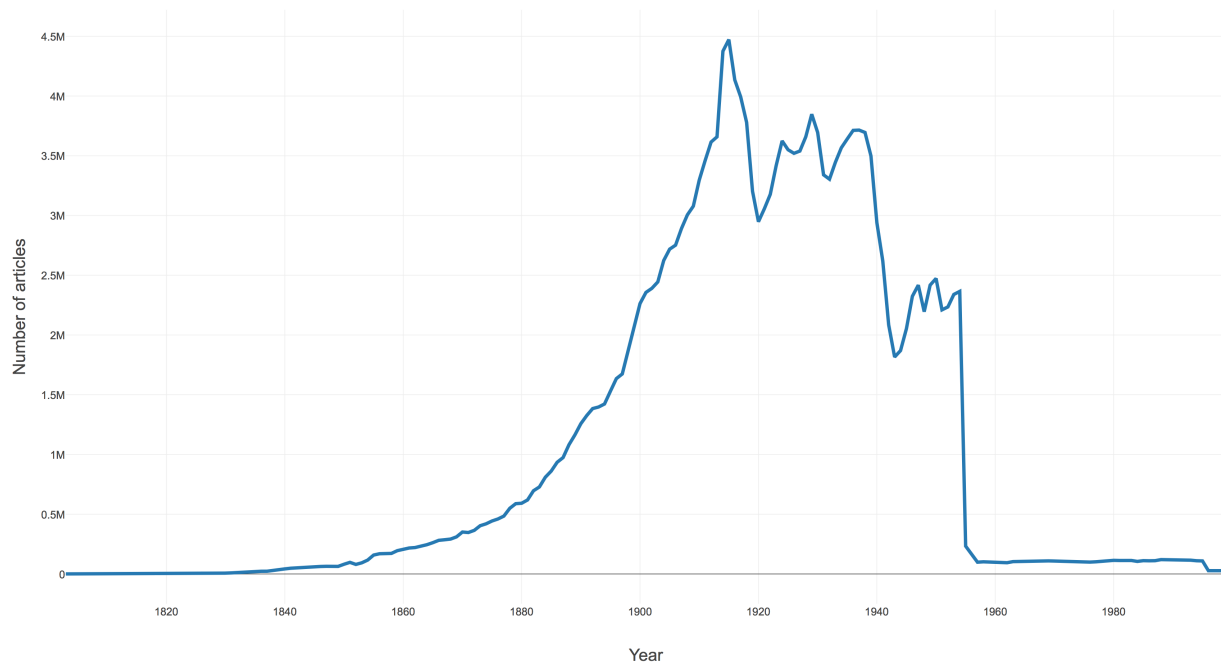
A dense grid of 100 newspaper clippings, each with a title and text, arranged in a 10x10 layout. The clippings are from various newspapers and cover a wide range of topics, including politics, social issues, and international relations. The titles are bold and prominent, while the text is in a smaller font. The overall appearance is that of a historical document or a collection of news items from a specific time period.



<https://easyzoom.com/image/139535>

There's a shift here from working with individual historical sources to analysing whole collections. This means extending our historical skill-set to include a range of new digital tools and methods. Once again, my [GLAM Workbench](#) provides a wide range of examples if you're interested in exploring the possibilities further.

But while digital resources like Trove challenge traditional research methods, they also need critical evaluation. As with any historical source, we have to ask questions about context, reliability, and completeness.



<https://plot.ly/dashboard/wragge:494/present>

Here's a chart that simply shows the number of digitised newspaper articles in Trove per year. What do you think caused the peak around 1914-15? Did something significant happen? Yes, it's because of the war, but not in the way you might think. The peak is a result of digitisation priorities – in the lead up to the centenary of World War I it was decided to focus on digitising newspapers from the war period. If digitisation continues (and that's dependent on continued funding), the peak will disappear. But for now it's a useful reminder that Trove itself is a historical construction. While we enjoy the wonderfully rich content it delivers to our browsers, we have to think about what's not there, what we can't find, and why.

Trove challenges us in other ways as well. Instead of simply using resources, we can build them. By adding tags or comments to records, by creating lists, or by correcting OCR results in newspapers, we can help other Trove users find what they're looking for. Through these sorts of contributions we are all building Trove together.

But we can also build completely new things based on the data Trove provides. Perhaps you'd like to play [Headline Roulette](#), a simple game that tests your historical knowledge by asking you to guess the publication date of a randomly selected newspaper article. Even better, you could use the ['do it yourself' version of Headline Roulette](#) to create your own customised game – limiting the newspaper articles to a particular topic or region. What about turning a Trove list into an online exhibition, or [building a Twitter bot](#) to share Trove collections?

Trove is many things. Most of all it's an opportunity to connect with Australia's past in new and interesting ways.

Selected resources

Using Trove

- [Trove](#)
- [Trove Tips & Tricks](#)
- [Explore Trove's digitised journals](#)
- [Locate Trove newspapers](#)

Trove as data

- [QueryPic](#)
- [GLAM Workbench](#)
- [Who belongs? Reading identity, ownership, and legitimacy.](#)

- [3,471 editorial cartoons from The Bulletin, 1886 to 1952](#)

Build and play

- [Headline Roulette](#)
- [Remix your own Headline Roulette](#)
- [A DIY Trove Exhibition](#)
- [Trove bots for all!](#)

Share

Tweet

LinkedIn

Reddit

Previous

← *Who belongs? Reading identity, ownership, and legitimacy*

Next

The multiplication of contexts →



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).