Discovering history at scale through Trove
The Australian Women’s Register

Helen Morgan
eScholarship Research Centre
The University of Melbourne
@HelenSnapper

www.womenaustralia.info @WomenAus #wmnhist

Australian Historical Association Annual Conference, 4 July 2018

Discovering History at Scale through Trove: the Australian Women’s Register
The Australian Women’s Register and Trove have a long association, going back to the Library’s pre-Trove experimental days.
We have an extensive, highly contextual data set underpinned by a relational informatic model,
that separates entities,
### Archival resources

**Australian National University Archives**
  - Details

  - Details

**National Archives of Australia, Sydney Office**
- [Personal papers of Prime Minister E G Whitlam](https://www.naa.gov.au/) Correspondence between E G Whitlam and Dr Thelma Hunter (Senior Lecturer in Political Science; School of General Studies) [Box 10], 1973. Series M621, Control Symbol 2016, Item Barcode 6834327. National Archives of Australia, Sydney Office.
  - Details

**National Library of Australia**
- Biographical cuttings on Thelma Hunter, containing one or more cuttings from newspapers or journals. Bib ID 1826932. National Library of Australia.
  - Details

**National Library of Australia, Manuscript Collection**
  - Details

  - Details

and resources about them,
and it has been freely, openly available online since the year 2000. Our data became the first external biographical source harvested into Trove.

So I begin with a caveat! The privilege and excitement of being first aside, I’m not sure how we stack up almost ten years down the track in terms of innovation, but my colleagues and I love Trove and this is how and why we use it.
The Register’s original remit was to document the location of archives relating to Australian women, and grew out of concerns of second-wave feminists and scholars about at-risk personal and non-government organisational records relating to women’s history. The purpose was and is to create and share feminist knowledge through the evidence of records and to create research infrastructure exposing feminist value in archival records for all those interested in the history of women in Australia.
As far as our relationship with Trove goes, we provide all our data, including relationship data and full descriptions; Trove harvest it and most of it ends up enriching entities,
in the People and Organisations zone, as well as forming the basis of new Trove entities, which in turn become nodes contextualising existing sources across the range of all the Trove zones.

I remember being queried some years ago by an archivist who wondered why we bothered anymore when so many archival repositories, for example, were now contributing record descriptions directly to Trove. Well, not all archival description is equal. Some is terrific, meaningful and fulsome. Some isn’t. Some description is contested. Some archivists are under the pump and think a minimal record is better than nothing and I agree. Most of us do what we can. Users have different views of records. In the Australian Women’s Register we can complement archival metadata through fuller description and/or, often the quicker route, by creating relationships between entities and records, entities and entities and so on. It is the power of the contextual and relational nature of our metadata that makes it worth doing, and sharing. Our data model supports an open view of data, and supports multiple viewpoints and relationships.
Muriel Heagney’s Trove record is an example containing biographical data from three contributors, including the Australian Women’s Register.
It also contains 5 relationships harvested from the Register – women and organisations, all of whom in turn have entries in Trove, providing a true exploration point for the user.

This is the basic way the Australian Women’s Register contributes to Trove. Not all our data is in there, but at last count we provide links back to Trove to more than two thirds of the 6000+ women and women’s organisations in the Register.
Many entries for women and women’s organisations only exist in Trove because they have been harvested from the Register, meaning a far broader range of women and women’s activities are being represented in the national context than previously.

Like many, we were greatly concerned when Trove and the National Library faced the funding crisis in 2016 that sparked the #fundTrove campaign. I knew we had some issues with the harvesting process and that it was only a matter of time before something went wrong and Trove staff would no longer have the capacity to troubleshoot and get things back on track. It worried me because we share data about living women, and we must be on top of the technology around its publication and sharing in order to be responsive to privacy and other concerns that arise.
The Register survives on a mixture of grant funding, small donations and the goodwill of volunteers and researchers willing to share their data. We have been fortunate to have been involved in several Australian Research Council funded projects along the way, leveraging the authority of the Register and our expertise in online data curation. These include Australian Women and Leadership, led by Patricia Grimshaw, Shurlee Swain and Judy Smart,
and, Trailblazing Women and the Law, led by Kim Rubenstein. The completed projects resulted in significant online publications, almost a thousand new biographical entries in the Register, easily as many bibliographical and archival entries, and 88 life history oral interviews undertaken for the National Library.
Both projects used the digitised newspapers zone of Trove extensively for similar purpose but with different outputs and evidence of use.

Why Trove digitised newspapers? I can’t imagine anyone here needs telling, but nevertheless I think it worth stating in relation to recovering women’s history and the stories of everyday people and families. You have to really work at recovering and excavating these histories.
In determining the values of a society, you need only investigate what gets discarded. Our literal and cultural detritus tell us as much about who we are as do our museums and libraries. Dumpster loads from our past would indicate that, in general, we value the new, the aesthetic, the whole, the extraordinary, the masculine, the Anglo, and the fast—not because our dumpsters are filled with these but rather because our textbooks are. On the **historical curb** rest the domestic, the broken, the consumable, the useful, the female, and the ordinary.


I’m going to quote a passage from Jennifer Sinor’s book *The Extraordinary Work of Ordinary Writing* here, because she says it so beautifully (above).

In these terms I think of Trove’s digitised newspaper zone as Australia’s historical curb. Historical newspapers were used extensively by researchers in the Women and Leadership project. Around 13% of the sources used to create biographical entries on women leaders were digitised newspaper articles and advertisements found in Trove, which may not sound significant but is much more than in the Register overall or in other individual Register projects. This reflects the editors’ decision to focus on women not already well documented, embrace broader definitions of leadership and to cover a wide variety of professional and other spheres of paid work, voluntary activities and different types of community engagement. The historical curb proved a good place to find them.
We entered the citations into our system and what you get in a typical entry alongside relationship data is biographical and bibliographical data. We fully cite the article with the url so users can go straight to the source.
Here is Lottie Lyell, an early Australian actor and director
Online Resources


and the sources used to write her entry,
including an advertisement for the film she starred in, *The Woman Suffers*, claimed to be Australia’s first feminist film.
This entry on nurse and trade unionist Gretta Lyons highlights many useful contemporary sources,
found on the historical curb, including one titled ‘Last Saturday's Storm’. I was intrigued by that, although it turns out to be the title only of the first item in a column of numerous untitled snippets.
Further in we find a notice that Williamstown Hospital was having trouble with its matrons, Gretta Lyons having resigned and her successor deciding that the pay was simply too poor to proceed. This was in 1906. The beauty – and it’s a simple beauty – of sending the reader away from our entry to the article in Trove is that word, context. We don’t have to spell it all out. In this case, directly preceding the observation about the matrons, which passes no judgement about the nurse who believed the pay wasn’t good enough,
we’re treated to a description of a Punch cartoon illustrating how crap Victorian cricketers were.
‘Females are shown in readiness to come to the rescue and take the place of the men.’ Hilarious.
And here’s that cartoon.

There were many more articles we could have given a curatorial imprimatur to, but entering data in the manual way that we were proved time consuming and beyond costly. We couldn’t keep it up, so, in the Register entries at least, ended up providing a link to the Trove person record, and letting Trove do the rest.
We took a different approach in the Trailblazing Women and the Law project, which was partly cost-cutting in terms of data entry.
Again, why not simply provide a link to the relevant person entry in Trove (which we know will have reasonable metadata because it has been harvested from us). This works well if indeed there is good metadata, but it can be hit and miss, especially around the issue of alternate forms of names.
So we decided, finally, to investigate tagging and harness the power of our dedicated casual researchers, who did way more work than we were able to pay them for. Much of this tagging work was done voluntarily, and I’d like to acknowledge Juliette Brodsky, who did the bulk of this work.

We set up a project-specific login that all project members had the login credentials for. I don’t think this is innovative, but it hadn’t occurred to us in the course of the Women and Leadership project, which I think was a shame and an opportunity missed. The idea was to game the Trove relevance and influence what results were returned first in the digitised newspaper zone.
We tagged articles with our internal project code and the woman’s name, sometimes multiple forms of the name but especially the fullest, as that appeared to reflect the logic of how Trove results were returned in relation to person records. All of these were public tags. Incidentally the project team corrected almost 23,000 lines of text during, and well beyond, the life of the project, at which point I must stop using ‘we’ and give that credit again to Juliette. Moving on from the project code and women’s name we tagged concepts, such as pioneering women lawyers, admitted to articles and early Jewish woman barrister.

Ultimately, to support the research and writing of biographical articles we identified many interesting and relevant newspaper articles that are now easier for everybody to find, because via the public tagging they are privileged for anyone searching Trove for information on women lawyers. That is as far as we got with the tagging. We were aiming to write some scripts to pull out articles based on our tagging for the online exhibition, but needs must as other exigencies took precedence, my coding skills were unpolished and it didn’t happen. The work wasn’t wasted though, as our main purpose of making public our research sources through influencing the results returned in Trove searches.
http://purl.org/au-research/grants/arc/LP120200367

We also decided to apply the PURL for our ARC project as a tag in order to create a public, reproducible, reportable data set of our research outcomes. At first we did this only to journal articles,
, published by the project team, and for the oral history interviews conducted for the Library’s Oral History and Folklore collection.
We wondered about the historic newspaper articles and I consulted Trove staff about this. Consensus was that it would be appropriate because it identifies our research data set. Everything we identified or used is tagged with the project code but the PURL was applied only to items of our own authorship and the historic newspaper articles we identified as relevant (that is, for example, not to oral history interviews with women lawyers conducted by other projects, which we tagged with only with the generic project code).
Nikki Henningham and Helen Morgan, 'The Australian Women's Register and the case of the missing apostrophe; or, how we learnt to stop worrying and love librarians', *The Australian Library Journal*, vol.65, no.3, 2016, pp.167-180

I finish with a plug for an article written by me and Australian Women’s Register co-editor Nikki Henningham, which has a strong focus on the relationship between the Register and Trove.

And back to my caveat – when you’re in research and the online space for the long haul you learn that not everything is achievable in that first flush of excited discovery, but the choices you make about sustainability and value mean you can future-proof to a certain extent and can scale up, with a nod to the conference theme, when the time is right!
The Australian Women’s Register

Co-editors

Dr Nikki Henningham
eScholarship Research Centre
The University of Melbourne
@NikkiHenningham

Helen Morgan
eScholarship Research Centre
The University of Melbourne
@HelenSnapper

www.womenaustralia.info @WomenAus #wmnhist

Thank you

Helen Morgan, Senior Research Fellow, eScholarship Research Centre, The University of Melbourne. Presented at the Australian Historical Association Annual Conference, 4 July 2018.
helen.morgan@unimelb.edu.au – http://www.helenmorgan.net/ – @HelenSnapper