

Fiction in 19th-century Australian Newspapers on Trove

Australian Historical Association panel Wednesday 4 July 2018; 11:00am-12:30pm



The project I want to talk about today seeks to discover and explore fiction in 19th-century Australian newspapers. These newspapers are of literary and historical interest because they were the main source of fiction for readers in this time and place, and the main avenue of publication for local authors; and hence, the main vehicles in the development of Australian literary culture.





This project is possible because Trove offers an excellent – and the world's largest – open-access, mass-digitised collection of historical newspapers. This digitised collection makes it possible to explore the fiction published in Australian newspapers for the first time in an extensive and systematic way.

TODAY . . .

- How I discovered & described fiction in *Trove's* mass-digitized newspapers
- 2. Some contributions to Australian literary and cultural history
- 3. Ongoing work to deploy literary data as a resource for scholarship, education and engagement

I'm planning to discuss three issues relating to this fortunate conjunction of literary historical circumstances and contemporary remediation:

- How I discovered fiction in these newspapers focusing on what I see as some general principles for large-scale historical research;
- Some of the contributions to literary and cultural history that these findings offer;
- Ongoing work to use large-scale data as the basis for scholarship as well as new collaborations and publics for historical research.



Here's a diagram of the process by which I discovered and described this fiction.

It includes some fairly standard features for research at scale.



It uses Trove's API or Application Programming Interface; which anyone can use to export data in machine-readable form.



The API stores the results in a database with an editable interface (and I'll demonstrate and discuss that interface in the final part of my talk).

I want to spend a bit more time with the two elements of this sequence that are particular to my project at the same time, as they have broader implications for historical research at scale.



The first of those elements is Step 1: Paratextual keyword query.

We know that we can search Trove newspapers using keywords to find articles of interest; by that approach, to discover fiction it might seem logical to search for authors or titles of fiction. But this approach has a number of significant problems:

- Have to progress author by author or title by title;
- You'd have to exclude a lot of irrelevant results obituaries, book reviews, literary gossip;
- More than simply being time-consuming, this approach would reinforce prior conceptions of the authors and titles published in Australian newspapers in that you'd find the ones you expected to be present;
- And it wouldn't accommodate the widespread anonymous publication of fiction or changes in titles as stories were published and republished.

The solution I developed I call the paratextual method. Takes advantage of the very high consistency of paratextual features in newspapers – the words used to introduce and describe articles. On this basis, it uses words such as "serial", "story", "novelist", "our author", and most usefully "chapter" to discover stories in the newspapers. The approach works because Trove zones pages into articles and defines and corrects title information, meaning that the paratextual information is both identified and more reliable than the rest of the OCR rendered text.

Because of the way the relevance ranking algorithm works to privilege articles where words appear in the title and are repeated throughout this approach meant that the vast majority of the initial thousands of results for searches with these different terms were fiction.

Take home for historical research at scale: methodologies need to be devised that combine disciplinary knowledge – in this case, of the generic and documentary context being investigated – with knowledge of the technical protocols, features and affordances of the digitised collection; either one without the other is insufficient.



The second element I want to highlight is Steps 5 and 8: Automatic, semi-automatic and manual exclusion and correcting of records.

Interaction of this paratextual method with features of Trove's digital collection made it possible to export fiction. But the harvested form is not especially useful for literary historical research.

Slide 10



This is an example of a record and you can see, for instance, that there's lots of different types of information in the "<heading>" field; but it's not structured in a way that is generalisable to other articles.

ARC enabled me to hire a bibliographer, Carol Hetherington, for 3 years to work on extensive manual processing of the harvested data to render it as useful literary historical – bibliographical – information.

```
Slide 11
```



This involved, for instance, extending the information in Trove's "<heading>" field into multiple separate categories including title, author, copyright and circulation information, chapter number and chapter title; and then descriptions of those categories, such as the form of authorial attribution.

Slide 12



It also involved incorporating additional information about fields, for instance, whether a newspaper was metropolitan, suburban or provincial; consistent location information; and stitching multiple different banner changes together under common newspaper titles.

Finally, there was extensive other information gleaned through bibliographical research (such as the common title or author name uniting the different titles and attributions in the newspapers; and details about gender and nationality of authors; sources of information and other sites of publication).



And after all that there was still an extensive process of editing and correcting the bibliographical data.

Take home for historical research at scale: creating data that is useful for humanities research is not a simple matter of sucking information out of Trove and analysing it for historical trends; prior to such analysis, significant research is often required to produce useful data; and this adds enormous value to what Trove currently makes accessible.

WHAT FICTION DID WE DISCOVER?

- Over 21,000 novels, novellas & short stories
- C19 newspapers (esp. 1865-1899):
 16,687 titles (incl. 7,399 short stories)
- Early C20 newspapers (1900-1914): -4,597 titles (incl. 1,008 short stories)
- 4,146 individual known authors
- 5,671 titles by unknown authors
- 194,108 text files



As a result of a harvest conducted in July 2015, when I estimate that around 30% of 19th century newspapers had been digitised, we discovered over 21,000 novels, novellas and short stories. We found that about 80% of the 19th century newspapers contained fiction – large amounts in most cases.

The numbers are not of the scale associated with so-called big-data research, but when you consider that, before this project, the bibliographical record for what we knew was published in 19th century Australian newspapers was in the hundreds this is a significant increase on existing knowledge.

SOME DISCOVERIES ABOUT FICTION IN COLONIAL AUSTRALIAN NEWSPAPERS



- Fiction from around the world (esp. Britain, Australia & America; but also Austria, Canada, France, Germany, Holland, Hungary, Italy, Japan, New Zealand, Russia, South Africa, Sweden)
- Hundreds of new Australian titles and authors
- Australian Australian fiction accorded more value than has been argued

Once we conducted analysis, we made discoveries that significantly transformed prior understandings of Australian literary and cultural history:

- Rather than these newspapers focusing largely or entirely on British fiction, as has been argued, this was a cosmopolitan literary and reading culture;
- There was much more Australian fiction than previous histories had proposed, with the discovered titles including hundreds of new titles by known Australian authors as well as hundreds of titles by over 60 previously unknown local authors (with the likelihood being that lots of the fiction where we've been unable to determine the authors' identity was Australian);
- There's evidence to suggest that this Australian writing was accorded significantly more value than has been argued, with the paratexts consistently highlighting local authorship, and there are some examples of how that occurs.

SOME DISCOVERIES ABOUT FICTION IN COLONIAL AUSTRALIAN NEWSPAPERS

- Consistent, prominent depiction of Aboriginal characters in local fiction
 - the "bush tradition" (but not as we've known it)
- New structure and organization for 19th-century Australian literary culture
 - cross-colonial provincial newspaper syndicates were *the* major publishers of Australian fiction and fiction in Australia

Topic80.txt

shepherd district homestead fence round kangaroo plains timber stations blacks tea timber night blacks tea timber started manuverandah started with the started started started with the started started started started with the started started started with the started started started started started started started started started of the started star

The local fiction also challenges a widespread argument in Australian postcolonial studies that 19th century fiction replicated the legal lie of terra nullius with Aboriginal people absent from Australian stories. In fact a characteristic feature of the Australian fiction was its consistent and prominently depiction of Aboriginal characters. The word cloud on the slide signifies the topic deemed by a machine learning method to be most strongly predictive of Australian writing – if you look closely you can see words indicative of the presence of Aboriginal characters (black, blacks, track, mob) alongside more rural settings, characters, and activities

The bush tradition lives – but not as we thought we knew it. Perhaps most remarkably, we discovered a new structure and organization for 19th-century literary culture: "bush tradition". Most previous histories have focused on books and metropolitan newspapers. In fact, provincial newspaper published more than either of their metropolitan counterparts, and did so in highly organised ways.

Exploring incidences of reprinting in the dataset shows that much of the provincial fiction was supplied by an extensive array of fiction syndication agencies, both Australian and international, that operated across colonial lines such that readers in outback New South Wales were reading the same stories as their Victorian, South Australian, Queensland, Tasmanian and sometimes their West Australian and New Zealand counterparts

This cross-colonial provincial culture challenges the view, in Alan Atkinson's words, that the Australian colonies operated as separate "nations" which "each of the Australasian capital cities ... the centre of its own world," including its own print culture. Many of these syndicates focused on Australian fiction and they emerge as the major publishers of such writing certainly in the 19th century and probably up until at least the 1970s.

The prevalence of local fiction in provincial newspapers from the late 1870s also dismantles a key tenet of Australian literary history, which locates the origins of local literary culture in the 1890s in the "Red Pages" of the Bulletin magazine; in fact, this Sydney-based publication followed in the footsteps of the provincial syndicates.

Let me move, now, beyond my findings for the project to consider how large-scale projects can enable other scholarship as well as new collaborations and publics for historical research. Largescale historical data takes such a long time (and often so much public money) to create and embeds so much important knowledge that I think it's essential we publish our data wherever possible.

The way in which we publish has implications for how it is used; and over time I've moved from what I now recognise as a broadcasting model of publishing to a collaborating one that not only makes data available but improves and extends it while facilitating sustainability of research findings.

```
Slide 17
```



All 21,000+ novels, novellas and short stories that we discovered are available in an editable database connected to Trove via the API (access it using that URL).

The database allows me to publish the data underpinning my arguments – so that others can explore, contest, confirm, or extend them.

In data-rich humanities research, data publication tends to involve putting that data in an openaccess repository, which has the benefit of the researcher not having to act as publisher, but the drawback of the data being effectively available only to those who have the computational skills to wrangle large datasets.

Aut	har	Select or enter a name.			•	Title	Enter	itie			
Ger	ider	Female \$ Nationalit	V Canada	n .		Newspaper	Enter	newspiper	OLong	th	
Sta		1628-01-01	End Date		1958-	2-01	Chapte	f Enter article topt	Curater		
Sho	wing 1-20 (Title ID	of 56 items. Publication Author	Author	Publication 1	litte	Common 1	Title	Search Exp	Start Date 1		root.
1	8519	Unatributed	Coghill, Annie L.	In Life and in Death: A Pag Family Histor	e cf	in Life and Page of fan History		The Herald (Fremantie, WA : 1867 - 1886)	1871-07-15	1871-07- 29	10
2	3022	Pienting, Mrs May Agres	Pleming, May Agnes	A Leap in the cr, Wedded B Wfe		A Leap in t or, Wed Yer Wite		The Matland Mercury and Hunter River General Advertiser (NSW : 1843 - 1893)	1873-11-22	1874-04- 25	01
3	18762	Flaming, Mrs May Agnes	Fleming, May Agnes	How the Lon Went Down	s Star	How the Lo Went Down		The Armidale Express and New England General Advertiser (NSW : 1856 - 1861; 1863 - 1889; 1891 - 1954)	1874-06-19	1874-06- 19	•1
4	14029	Unathibuted	Fleming, May Agnes	Leoline; or, TI Mask, the Ca and the King		The Midnig Queen	ht	The Newcastle Chronicle (NSW : 1866 - 1876	1875-02-16	1875-06- 19	01
9	3557	Unatributed	Pieming, May Agnes	A Profound Mystery		A Terrible 5	Hecret	Kerang Times and Swan Hill Gazette (Vic. : 1877 - 1889)	1868-05-29	1869-12- 14	01
6	23290	Cleary, Kate M.	Cieary; Kate M.	A July Joke		A July Joka	•	The Dawn (Sydney, NSW : 1868 - 1905)	1893-03-04	1893-03- 04	01
7	3116	Unatributed	Fleming, May Agnes	Paulina Lale		A Leap in 1 or, Wed Yet Wife		The Matland Mercury and Hunter River General Advertiser (NDW : 1040 - 1090)	1893-07-11	1893-11- 21	10
8	14148	Unatributed	Fleming, May Agnes	Srowball		Lost for a V	Noman	The Maitland Daily Mercury (NSW : 1894 - 1939)	1894-02-22	1894-05-	

In a database, it offers a resource for research, whether it is computational or non-computational; there are various facets to limit searches as well as keyword searching, and if you've got good eyes you might be able to see that I've limited the search here to Canadian women writers – and we can see that romance writer May Agnes Fleming is very well represented in that category.

Con Trev Con Trev	nceston Examiner (1 emon Title : Half a C e Title : Half a Cent emon Name : Nottage, J e Name : Nottage, J	entury Ago ary Ago jo, John Silvester John								
	chapter Number		Chapter Url	Day	Cate	Month	Year	Page Number	Article ID	
F	Crapon narion				Conte		Para I			
1	1	OVER THEIR CUPS OF TEA.	http://nia.gov.au/nia.news- article00258064	Saturday	8	05	1680	0	38258064	0
2	н	FPWOOD, AND ITS INMATES.	http://itia.gov.su/ria.news- artick08258259	Saturday	15	05	1680	0	38258259	۲
з	н.,	MR. MAUVAIN AT HOME	htp://ria.gov.su/nia.news- artck/36206465	Esturday	22	05	1680	0	38258465	•
4	v	AN UNLOCKED-FOR VISITOR	http://tila.gov.du/hla.news- article/38258992	Saturday	5	06	1880	0	38258962	۲
5	v	AT THE WHITE HOUSE.	http://irla.gov.au/rla.news- artciw38259027	Seturday	12	06	1880	0	38259227	۲
6	VI	VAL FERRET'S CLIENT.	http://inla.gov.au/nla.news- article38259472	Schurdby	19	06	1680	0	38259472	۲
7	VIE	LIZZE WARREN	http://ria.gov.ku/nla.news- article/38260725	Saturday	25	08	1880	0	38259725	۰
8	1X	A RENCONTRE.	http://nla.gov.au/nla.news- antce/05/21/9944	Seturday	3	07	1680	0	38259964	•
9	×	THE SPECTRE AT THE WINDOW	http://inia.gov.au/nia.news- article36200225	Saturday	10	07	1680	0	38260225	•
10	х	FERRET MAKES A MOVE.	http://inia.gov.au/ma.news- article/38260464	Seturday	17	07	1880	0	38260464	۲
11	XI	THE FUGITIVE AT FIRMOOD.	http://www.gov.au/nla.news- artice/28280/19	Seturday	24	07	1600	0	36260719	۲
12	XE	ELUSTRATIVE OF SHAKESPEARE'S ASSERTION THAT THE COURSE OF	http://kia.goz.su/kia.news- article/202000	Saturday	31	07	1880	0	38260960	•

You can explore the individual instalments or chapters of the fiction. You can also export textual and bibliographical data for computational analysis, whether of the whole database or a defined subset.

These stories could certainly provide a resource for traditional research and general publications.





There's my book, on the left, based on this collection and on the right is a collection I edited with Canberra's own Obiter Press last year publishing five newly discovered stories by important 19th century Australian feminist, socialist and novelist Catherine Martin. This year we're working, with different editors, on two further collections of 19th century Australian Xmas stories and mining fiction.

However, I am also exploring how data publication can enable new forms of collaboration beyond the academic and publishing spheres. Because the database is connected to Trove's API, members of the public can find a story they're interested in in the database...



... correct the text in Trove and ...

See chapter	In newspaper Update text from trove Edit chapter details Delete
Chapter Number	x
Chapter Title	THE GREAT MAIL ROBBERY.
Chapter Url	http://nla.gov.au/nla.news-article857881
Full Date	1882-09-16
Page Number	2
Corrections	1
Word Count	5975
Illustrated	N
Last Corrected	2018-02-27
Newspaper Title	The Maitland Mercury and Hunter River General Advertiser (NSW : 1843 - 1893)
Trove Title	Brought Home
Chapter Text	FICTION. (From English, American, and other Periodicals.) BROUGHT HOME. CHAPTER X THE GREAT MAIL BOBBKBY. After having hie hiding-place behind the bookcase discovered in euch an unceremonious fashion, and almoBt miraculeusly escaping the builtet of the vii- lain ForbeB, Belmont Brown, the post-office detec- tive, idd not spare time or pains to make himmelf scarce in the locality. Being by nature n cautious and far-sighted perbon (else be would never how been a success as a detec- tive, he bad

... draw that corrected text back into the database, thus improving the accuracy of keyword searching in both sites.

Vie	w Title Add Chapte	r Export Text							
he	Queenston r (Brisbar mon T/ 2 Hundred e Tit/ Hundred Pou mv ame : Stephens	unds , J. Brunton							
	Chapter Number 1	Chapter Title	Day	Date	Month	Year	Page Number	Article ID	
1	1	HE AND SHE.	Saturday	19	02	1876	9	18340852	۲
2		A RESPECTABLE MAN'S THOUGHTS.	Saturday	26	02	1876	9	18340981	۲
3		SANDIE'S MEANNESS.	Saturday	4	03	1876	9	18341113	۲
4	IV	BRAITHWAITE'S VISITOR.	Saturday	11	03	1876	9	18341228	۲
5	v	AUTHORSHIP	Saturday	18	03	1876	9	18341363	0
6	VII	A MONTH AFTER.	Saturday	25	03	1876	9	18341487	۲
7	IX	None	Saturday	1	04	1876	9	18341605	۲
8	x	A DARK NIGHT'S WORK.	Saturday	8	04	1876	9	18341732	۲
9	хі	NOT YET.	Saturday	15	04	1876	9	18341858	۲

They can also add missing chapters or instalments that they discover in the digitized newspapers, or entirely new titles as well as to...

Author		Select or enter a name		• TI	tie	Enter title	8				
Gen	nder	+ Nationality			t Ne	ewspaper	Enter news	paper	OLer	ngth	4
Star		1828-01-01	End Date		1958-12-	31	Chapter	Enter article text	Curat		
#	wing 1-20 Title IJ	546 items. Publication Author	Author	Publication	Title	Common	Title	Newspaper Title	Start Date	End Date	
#	1000		Author	Publication	Title	Common	Title	Newspaper Title	Start Date	End Date	
#	1000		Author	Publication	Title	Common	Title	Newspaper Title	Start Date	End Date	
# 1	1000		Author Dickens, Charles	Publication The Mystery Edwin Droo	/ of		Title try of Edwin	Newspaper Title Australian Town and Country Jourr (Sydney, NSW : 1870 - 1907)		End Date	01
1 2	Title IJ	Publication Author	Dickens,	The Myster	/ of d	The Myste Drood		Australian Town and Country Journ	al 1871-03-04		01 01
•	Title IJ	Publication Author	Dickens, Charles Dickens,	The Mystery Edwin Droo	/ of d d ns in	The Myste Drood The Myste Drood	ry of Edwin ry of Edwin Recollections	Australian Town and Country Journ (Sydney, NSW : 1870 - 1907) Evening News (Sydney, NSW : 186	al 1871-03-04	1871-06-03	1

...publish the completed and corrected stories.

Since the database was launched in March this year I've had over 400 members of the public sign up to the research community: many focus on text correction but a substantial number add new stories to the database, often by focusing on particular authors or previously undigitized newspapers. Last time I checked over 300 new titles had been added and many more than that had been corrected.

Trov	Maitland Mercury and mon Title : The Forsak e Title : Abandoned mon Name : Sheldon, e Name : Unattributed	en Bride	aa)						
Sho	wing 1-18 of 18 items.	>							
٠	Chapter Number 1	Chapter Title	Day	Date	Month	Year	Page Number	Article ID	
1	HI	A WIFE'S MALEDICTION. THE BITTERNESS OF DEATH. SHE WILL LIVE, PLEASE GOD.	Saturday	21	08	1880	4	811476	
2	TV - VI	I HAVE BEEN ROBBED: A STRUGGLE FOR LIFE: LADY LEICESTER	Saturday	28	08	1880	4	811656	۲
3	VII-IX	A SINGULAR ADVENTURE: "I TELL YOU I AM NOT YOUR COUSIN": MADAM LEICESTER'S RECEPTION	Saturday	4	09	1880	4	811797	0
4	X - XII	THE STAR OF THE BALL: A BLISSFUL MOMENT: REVENGE	Saturday	11	09	1880	4	811940	۲
5	XIII-XV	MISTRESS OF THE FIELD: BAFFLED AT EVERY TURN: THEY WILL DIE GAME	Saturday	18	09	1880	4	812117	•
6	XVI - XVIII	NONE: I SAW HER MEET HIM THERE: DID SHE DENY IT?	Saturday	25	09	1880	4	812285	۲
7	XIX - XX	CAROLINE'S STRATEGY: AN UNREASONABLE DEMAND	Saturday	2	10	1880	4	812394	۲
8	200 -2003	I SHALL CONQUER IN THE END! : THE EARL'S FEARFUL DREAMS :	Saturday	9	10	1880	4:	812594	۲
9	XXIII - XXV	KENNETH MALCOLM'S LETTER : PREPARING FOR THE DWORCE: IT SHALL NEVER BE	Saturday	16	10	1880	4	812750	۲

A further collaboration is with Trove itself. I worked with the Trove team to link our two databases so that they could harvest all of these records of fiction and represent them in Trove.

Now, when someone searches the National Library of Australia's catalogue ...

Slide 26



They'll find records of fiction in Australian newspapers including stories where Trove indexes other versions: here book editions...



As well as stories where the only record in Trove is from the digitised newspapers.

Slide 28

1904-1905, Eng magazine, othe The Lone Inn	glish, Periodical, Journal, r edition:	Get this edition
Hume, Fergus		
Bookmark: https://trove.r Serial fiction 19 installments.	Na.gov.au/vension/252883357	
The Western Champion a : 1892 - 1922), 1904-12-2	nd General Advertiser for the Central-Western Districts (Barcaldine, Qld. 6 - 1905-03-27	
English		
View all editions	rev Next D edition 2 of 3	
Edition details		
Cite this		
Title	The Lone Inn	
Creator	Hume, Fergus	
Published	The Western Champion and General Advertiser for the Central-Western Districts - 1905-03-27	(Barcaldine, Qld. : 18
Physical Description	Serial fiction 19 instaliments.	
Part Of	Australian Newspaper Fiction Database	
Language	English	
Related Resource	Read chupter I: THE ADVENTURE AT THE INN. Published Monday 1904-12-28 Read chupter I: Published Monday1905-01-902 Read chupter II: Published Monday1905-01-902 Read chupter III: Published Monday1905-01-92 Read chupter IV: Published Monday1905-01-92 Read chupter V: Published Monday1905-01-30 page 2 Read chupter V: Published Monday1905-03-30 Read chupter V: Published Monday1905-03-27 Read chupter V: Published Mo	

Clicking on harvested records shows the detail, including links to the digitized newspaper articles that comprise the instalments of the stories and to the Australian Newspaper Fiction Database. This connection with Trove completes what we might call the humanities data cycle, in that the project both draws on and feeds back into a national digital collection, thus enriching knowledge and accessibility of the documentary record, while also ensuring the sustainability of research findings.

In collaboration with some colleagues in education I'm also hoping to develop a project wherein this data publication can act as a resource for teaching in subject English in Australian secondary schools that seeks:

- to bring the history of reading (of Australian literature and literature in Australia) into conversation with new media literacies (such as the nature of OCR, how collections are created); and
- enables students to act as independent researchers and curators, while engaging in the discussion of literary works that has typically comprised the treatment of literature

I hope what I've shown today are some of the ways in which large-scale research in the humanities is draws on existing practices and knowledge – relating to the historical context we're exploring, the types of arguments relevant to our fields, and existing collaborations of academics and publishers – but it also enables new practices and knowledge – including new collaborations – that have the potential to revise our understandings and connect us to new partners and publics.