hacking heritage

exploring the limits of access

Tim Sherratt · @wragge · timsherratt.org
these slides

https://slides.com/wragge/hacking-heritage-openglamnow
A hack can be elegant or kludgy, authored from scratch or patched together and remixed—the important thing is getting things done, pushing the boundaries of what the humanities can do, what effects it can have in the world, and where.

— Mark J Olson, 'Hacking the humanities: Twenty-first-Century literacies and the ‘becoming other’ of the humanities'
access is constructed
Explore Trove’s digitised journals

112 journals and 86,211 articles added since 4 July 2019

There's lots of exciting new digitised content being added to Trove's journals zone, but it's not always easy to find and search.

This page lists journals that have been digitised by the NLA and have searchable records for individual articles. This means you can search inside the journal, just like you do in the newspapers zone.

To search inside these journals, just click on the titles you're interested in below, then enter keywords in the search box.

Enter your search keywords

No titles selected.

Search Trove

Clear everything
• what's available?
• is it in the form I need?
• do I have the technical skills?
• what does it really mean?
what's available?

*I don't know where to start...*
Explore Trove's digitised journals

112 journals and 86,211 articles added since 4 July 2019

There's lots of exciting new digitised content being added to Trove's journals zone, but it's not always easy to find and search.

This page lists journals that have been digitised by the NLA and have searchable records for individual articles. This means you can search inside the journal, just like you do in the newspapers zone.

To search inside these journals, just click on the titles you're interested in below, then enter keywords in the search box.

Enter your search keywords

No titles selected.

Search Trove

Clear everything
some random hacks

Get a random illustrated advertisement from the Australian Womens Weekly

In [9]:
get_random_article(title='112', illustrated='true', category='Advertising')

Out[9]:
{'id': '53250867',
'url': '/newspaper/53250867',
'heading': 'Advertising',
'category': 'Advertising',
'title': {'id': '112',
'value': 'The Australian Women’s Weekly (1933 - 1982)'
},
'date': '1980-06-18',
'page': 101,
'pageSequence': 101,
'relevance': {'score': '0.24747185', 'value': 'may have relevance'},
'troveUrl': 'https://trove.nla.gov.au/ndp/article/53250867?searchTerm=%22had%22'}

Get a random cartoon

In [10]:
get_random_article(illtype='Cartoon')

Out[10]:
{'id': '75198942',
'url': '/newspaper/75198942',
'heading': 'Don’t Tell Auntie An Hilarious Christmas-time Episode',
'category': 'Article',
'title': {'id': '251', 'value': 'Sunshine Advocate (Vic.: 1924 - 1954)'},
'date': '1937-12-17',
'page': 1,
'pageSequence': '1 S',
'relevance': {'score': '1.9126042', 'value': 'likely to be relevant'},
'snippet': 'George Cunningham was not carrying his auntie’s luggage; he was only directing the efforts of all the

Get a random article from 1930

In [11]:
get_random_article(year='1930')

Out[11]:
{'id': '113686939',
'url': '/newspaper/113686939',
'heading': 'DID BUILDER ERR? President Townsend Outspoken',
'category': 'Article',
'title': {'id': '410',
'value': 'Gilgandra Weekly and Castlereagh (NSW : 1929 - 1942)'
},
'date': '1930-11-20',
'page': 7,
'pageSequence': 7,
'relevance': {'score': '0.9030686', 'value': 'likely to be relevant'},
'snippet': 'In this case it appears that the Council has deliberately flouted,” declared the President (Cr.
'troveUrl': 'https://trove.nla.gov.au/ndp/article/113686939?searchTerm=%22not%22'}
Trove News Bot

Followed by
- USQ Centre for Heritage and Culture
- History Unearthed
- 278 others you follow

Trove News Bot @TroveNewsBot · 2h
Found! 14 Aug 1878: Tasmanian Telegrams. TORQUAY, TUESDAY EVENING; Devon Herald, nla.gov.au/nla.news-artic...

Trove News Bot @TroveNewsBot · 3h
lists can be useful too!

GLAM datasets harvested from data.gov.au (8 July 2019)

See GLAM Workbench for harvesting code and summary.

- Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS)
- Australian Museum
- History Trust of South Australia
- Libraries Tasmania
- Mount Gambier Library
- Museum of Applied Arts and Sciences
- NSW State Archives
- National Archives of Australia
- National Library of Australia
- National Portrait Gallery
- PROV Public Record Office
- Queensland State Archives
- South Australian Museum
- State Library of NSW
- State Library of New South Wales
- State Library of Queensland
- State Library of South Australia
- State Library of Victoria
- State Library of Western Australia
- State Records
- State Records Office of Western Australia
- Tasmanian Museum and Art Gallery
- Western Australian Museum

<table>
<thead>
<tr>
<th>Creative Commons Attribution</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creative Commons Attribution 3.0 Australia</td>
<td>244</td>
</tr>
<tr>
<td>Creative Commons Attribution 4.0</td>
<td>237</td>
</tr>
<tr>
<td>Creative Commons Attribution 4.0 International</td>
<td>146</td>
</tr>
<tr>
<td>Creative Commons Attribution 2.5 Australia</td>
<td>32</td>
</tr>
<tr>
<td>Creative Commons Attribution-NonCommercial</td>
<td>10</td>
</tr>
<tr>
<td>Other (Open)</td>
<td>5</td>
</tr>
<tr>
<td>notspecified</td>
<td>5</td>
</tr>
<tr>
<td>Creative Commons Attribution Share-Alike 4.0</td>
<td>3</td>
</tr>
<tr>
<td>Creative Commons Attribution 3.0</td>
<td>3</td>
</tr>
<tr>
<td>Creative Commons Attribution Non-Commercial 4.0</td>
<td>2</td>
</tr>
<tr>
<td>Custom (Other)</td>
<td>1</td>
</tr>
</tbody>
</table>
but what's inside?

GLAM CSV Explorer

Cultural institutions are making collection data available as machine-readable downloads. But how can researchers explore the shape and meaning of this data? How do they know what types of questions they can ask?

This notebook provides a quick overview of CSV-formatted data files, particularly those created by GLAM institutions (galleries, libraries, archives, and museums).

It looks at each column in the dataset and tries to identify the type of data inside. It then attempts to tell you something useful about it. There’s some more details about the process below.

Given all the possible variations in recording and formatting data, there will be errors. But hopefully this will provide you with a useful starting point for further exploration.

More information

The GLAM CSV Explorer is a Jupyter notebook, combining live Python code with text and widgets in a form that's easy to hack and build upon. If you can't see any code at the moment, you're probably in 'app mode'. Just click on the Edit app button at the top of the screen to see what's going on underneath.

The app makes heavy use of Pandas, the all-purpose toolkit for working with tabular data. Pandas is quick and powerful, but has so many options it can be difficult to know where to start. You might like to poke around in the code for ideas.

To analyse a CSV, the explorer looks at things like the datatype of a column, and the number of unique values it holds. It also applies a variety of regular expressions to look for dates and numbers. Depending on what it finds, it extracts some summary information, and tries to visualise the results using WordCloud and Altair.

To-do

- Handle geo-spatial data
- Display images from urls

https://glam-workbench.github.io/csv-explorer/
is it in the form I need?

downloads are cool but...
spoken in Parliament
Commonwealth of Australia Hansard

These are Hansard XML files downloaded from the Parliament of Australia website by Tim Sherratt. The original harvest was run in 2016. A complete reharvest was undertaken in April 2019. Some of the file names have changed between harvests.

The files are made available on the Parliament website under a CC-BY-NC-ND licence.

Currently this repository includes files for the House of Representatives and the Senate from 1901 to 1980. The Hansard file format changes after 1980.

Browse the files or download a zip file of the lot (it's big). For convenience there's also individual zip files for each year.

More info

- Commonwealth Parliamentary Debates (Hansard) – Jupyter notebooks to harvest and explore XML formatted versions of Hansard.
- Historic Hansard in my Research Notebook
- Historic Hansard – my own version of Hansard, optimised for easy browsing.
- Documentation of the Historic Hansard site.

https://github.com/wragge/hansard-xml
one page per day

Historic Hansard

Commonwealth of Australia parliamentary debates presented in an easy-to-read format for historians and other lovers of political speech.

House of Representatives, 1901 to 1980  Senate, 1901 to 1980

Built by Tim Sherratt.  
Support this project on Patreon.

Hansard is licensed for reuse by the Parliament of Australia under a CC-BY-NC-ND licence.
3,471 Bulletin editorial cartoons

https://glam-workbench.github.io/trove-journals/
Previously this kind of research would have taken months but now thanks to Tim’s shortcut can be done in a matter of days.

— Guy Hansen, National Library of Australia
scraping data from the NAA

<table>
<thead>
<tr>
<th>series</th>
<th>total_items</th>
<th>date_from</th>
<th>date_to</th>
<th>Open</th>
<th>OWE</th>
<th>NYE</th>
<th>Closed</th>
<th>digitised_files</th>
<th>digitised_pages</th>
<th>% open</th>
<th>% digitised</th>
</tr>
</thead>
<tbody>
<tr>
<td>B13</td>
<td>20,194</td>
<td>1800</td>
<td>2005</td>
<td>19,786</td>
<td>8</td>
<td>400</td>
<td>0</td>
<td>354</td>
<td>5,043</td>
<td>97.98%</td>
<td>1.75%</td>
</tr>
<tr>
<td>B6003</td>
<td>3</td>
<td>1904</td>
<td>1959</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>BP343/15</td>
<td>2,571</td>
<td>1916</td>
<td>1955</td>
<td>2,566</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>85</td>
<td>176</td>
<td>99.81%</td>
<td>3.31%</td>
</tr>
<tr>
<td>D2860</td>
<td>1</td>
<td>1902</td>
<td>1957</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>D5036</td>
<td>1</td>
<td>1906</td>
<td>1935</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>D596</td>
<td>11,395</td>
<td>1871</td>
<td>1971</td>
<td>2,983</td>
<td>31</td>
<td>8,381</td>
<td>0</td>
<td>185</td>
<td>3,031</td>
<td>26.18%</td>
<td>1.62%</td>
</tr>
<tr>
<td>E752</td>
<td>722</td>
<td>1905</td>
<td>1941</td>
<td>719</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>717</td>
<td>9,310</td>
<td>99.58%</td>
<td>99.31%</td>
</tr>
<tr>
<td>J2481</td>
<td>858</td>
<td>1897</td>
<td>1903</td>
<td>858</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>858</td>
<td>2,031</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>J2482</td>
<td>799</td>
<td>1902</td>
<td>1912</td>
<td>799</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>798</td>
<td>3,153</td>
<td>100.00%</td>
<td>99.87%</td>
</tr>
<tr>
<td>J2483</td>
<td>14,438</td>
<td>1903</td>
<td>1956</td>
<td>14,436</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>14,436</td>
<td>79,210</td>
<td>99.99%</td>
<td>99.99%</td>
</tr>
<tr>
<td>J3115</td>
<td>161</td>
<td>1899</td>
<td>1928</td>
<td>161</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>161</td>
<td>1,344</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>K1145</td>
<td>4,816</td>
<td>1900</td>
<td>1955</td>
<td>4,791</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>175</td>
<td>874</td>
<td>99.48%</td>
<td>3.63%</td>
</tr>
<tr>
<td>P437</td>
<td>4,958</td>
<td>1901</td>
<td>1940</td>
<td>4,945</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>442</td>
<td>99.74%</td>
<td>0.36%</td>
</tr>
<tr>
<td>P526</td>
<td>2</td>
<td>1909</td>
<td>1918</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>50.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>PP4/2</td>
<td>613</td>
<td>1903</td>
<td>1947</td>
<td>610</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>28</td>
<td>1,512</td>
<td>99.51%</td>
<td>4.57%</td>
</tr>
<tr>
<td>PP6/1</td>
<td>6,010</td>
<td>1906</td>
<td>1978</td>
<td>1,863</td>
<td>33</td>
<td>4,109</td>
<td>5</td>
<td>245</td>
<td>6,461</td>
<td>31.00%</td>
<td>4.08%</td>
</tr>
<tr>
<td>SP11/26</td>
<td>27</td>
<td>1902</td>
<td>1902</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>84</td>
<td>100.00%</td>
<td>18.52%</td>
</tr>
<tr>
<td>SP11/6</td>
<td>191</td>
<td>1902</td>
<td>1947</td>
<td>101</td>
<td>0</td>
<td>90</td>
<td>0</td>
<td>1</td>
<td>323</td>
<td>52.88%</td>
<td>0.52%</td>
</tr>
<tr>
<td>SP115/1</td>
<td>1,787</td>
<td>1884</td>
<td>1943</td>
<td>1,787</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>285</td>
<td>100.00%</td>
<td>0.50%</td>
</tr>
<tr>
<td>SP115/10</td>
<td>6</td>
<td>1884</td>
<td>1888</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>SP42/1</td>
<td>16,256</td>
<td>1881</td>
<td>1960</td>
<td>15,525</td>
<td>0</td>
<td>731</td>
<td>0</td>
<td>3,253</td>
<td>45,862</td>
<td>95.50%</td>
<td>20.01%</td>
</tr>
<tr>
<td>SP726/1</td>
<td>6</td>
<td>1902</td>
<td>1959</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>ST84/1</td>
<td>2,765</td>
<td>1855</td>
<td>1975</td>
<td>2,758</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>434</td>
<td>13,979</td>
<td>99.75%</td>
<td>15.70%</td>
</tr>
</tbody>
</table>
The Real Face of White Australia

Join us in transcribing records that document the lives of ordinary people living under the restrictions of the White Australia Policy.

GET STARTED!

Closed access

Overview

Under the Australian Archives Act 1983 most Commonwealth records are opened to public scrutiny after twenty years (this was reduced from thirty years in 2010). But the Act also defines ‘exempt’ records that can be withheld from the public for a variety of reasons, including the defence of national security, and the protection of individual privacy. Access under the Act is not an inevitable destination, but a process that may result in records with the access status of ‘closed’.

Here you can explore these closed files. Why can’t we look at them? How old are they? What are we really being protected against?

14370

reasons

on 1 January 2016

why these files have been closed

28

years

is the average age of these files based on their date of their earliest content

58

series

contain closed files

1128

redacted:

note of exempt, exempted and restricted

accessed from 13 Oct 2020 by

buburt

who is the keeper of public records in the practice of state archives.

The Real Face of

White Australia

THE REAL FACE OF

WHITE AUSTRALIA

GET STARTED!

Overview

Under the Australian Archives Act 1983 most Commonwealth records are opened to public scrutiny after twenty years (this was reduced from thirty years in 2010). But the Act also defines ‘exempt’ records that can be withheld from the public for a variety of reasons, including the defence of national security, and the protection of individual privacy. Access under the Act is not an inevitable destination, but a process that may result in records with the access status of ‘closed’.

Here you can explore these closed files. Why can’t we look at them? How old are they? What are we really being protected against?

14370

reasons

on 1 January 2016

why these files have been closed

28

years

is the average age of these files based on their date of their earliest content

58

series

contain closed files

1128

redacted:

note of exempt, exempted and restricted

accessed from 13 Oct 2020 by

buburt

who is the keeper of public records in the practice of state archives.

The Real Face of

White Australia

THE REAL FACE OF

WHITE AUSTRALIA

GET STARTED!

Overview

Under the Australian Archives Act 1983 most Commonwealth records are opened to public scrutiny after twenty years (this was reduced from thirty years in 2010). But the Act also defines ‘exempt’ records that can be withheld from the public for a variety of reasons, including the defence of national security, and the protection of individual privacy. Access under the Act is not an inevitable destination, but a process that may result in records with the access status of ‘closed’.

Here you can explore these closed files. Why can’t we look at them? How old are they? What are we really being protected against?

14370

reasons

on 1 January 2016

why these files have been closed

28

years

is the average age of these files based on their date of their earliest content

58

series

contain closed files

1128

redacted:

note of exempt, exempted and restricted

accessed from 13 Oct 2020 by

buburt

who is the keeper of public records in the practice of state archives.

The Real Face of

White Australia

THE REAL FACE OF

WHITE AUSTRALIA

GET STARTED!

Overview

Under the Australian Archives Act 1983 most Commonwealth records are opened to public scrutiny after twenty years (this was reduced from thirty years in 2010). But the Act also defines ‘exempt’ records that can be withheld from the public for a variety of reasons, including the defence of national security, and the protection of individual privacy. Access under the Act is not an inevitable destination, but a process that may result in records with the access status of ‘closed’.

Here you can explore these closed files. Why can’t we look at them? How old are they? What are we really being protected against?

14370

reasons

on 1 January 2016

why these files have been closed

28

years

is the average age of these files based on their date of their earliest content

58

series

contain closed files

1128

redacted:

note of exempt, exempted and restricted

accessed from 13 Oct 2020 by

buburt

who is the keeper of public records in the practice of state archives.

The Real Face of

White Australia

THE REAL FACE OF

WHITE AUSTRALIA

GET STARTED!

Overview

Under the Australian Archives Act 1983 most Commonwealth records are opened to public scrutiny after twenty years (this was reduced from thirty years in 2010). But the Act also defines ‘exempt’ records that can be withheld from the public for a variety of reasons, including the defence of national security, and the protection of individual privacy. Access under the Act is not an inevitable destination, but a process that may result in records with the access status of ‘closed’.

Here you can explore these closed files. Why can’t we look at them? How old are they? What are we really being protected against?

14370

reasons

on 1 January 2016

why these files have been closed

28

years

is the average age of these files based on their date of their earliest content

58

series

contain closed files

1128

redacted:

note of exempt, exempted and restricted

accessed from 13 Oct 2020 by

buburt

who is the keeper of public records in the practice of state archives.

The Real Face of

White Australia

THE REAL FACE OF

WHITE AUSTRALIA

GET STARTED!

Overview

Under the Australian Archives Act 1983 most Commonwealth records are opened to public scrutiny after twenty years (this was reduced from thirty years in 2010). But the Act also defines ‘exempt’ records that can be withheld from the public for a variety of reasons, including the defence of national security, and the protection of individual privacy. Access under the Act is not an inevitable destination, but a process that may result in records with the access status of ‘closed’.

Here you can explore these closed files. Why can’t we look at them? How old are they? What are we really being protected against?

14370

reasons

on 1 January 2016

why these files have been closed

28

years

is the average age of these files based on their date of their earliest content

58

series

contain closed files

1128

redacted:

note of exempt, exempted and restricted

accessed from 13 Oct 2020 by

buburt

who is the keeper of public records in the practice of state archives.

The Real Face of

White Australia

THE REAL FACE OF

WHITE AUSTRALIA

GET STARTED!

Overview

Under the Australian Archives Act 1983 most Commonwealth records are opened to public scrutiny after twenty years (this was reduced from thirty years in 2010). But the Act also defines ‘exempt’ records that can be withheld from the public for a variety of reasons, including the defence of national security, and the protection of individual privacy. Access under the Act is not an inevitable destination, but a process that may result in records with the access status of ‘closed’.

Here you can explore these closed files. Why can’t we look at them? How old are they? What are we really being protected against?

14370

reasons

on 1 January 2016

why these files have been closed

28

years

is the average age of these files based on their date of their earliest content

58

series

contain closed files

1128

redacted:

note of exempt, exempted and restricted

accessed from 13 Oct 2020 by

buburt

who is the keeper of public records in the practice of state archives.

The Real Face of

White Australia

THE REAL FACE OF

WHITE AUSTRALIA

GET STARTED!

Overview

Under the Australian Archives Act 1983 most Commonwealth records are opened to public scrutiny after twenty years (this was reduced from thirty years in 2010). But the Act also defines ‘exempt’ records that can be withheld from the public for a variety of reasons, including the defence of national security, and the protection of individual privacy. Access under the Act is not an inevitable destination, but a process that may result in records with the access status of ‘closed’.

Here you can explore these closed files. Why can’t we look at them? How old are they? What are we really being protected against?

14370

reasons

on 1 January 2016

why these files have been closed

28

years

is the average age of these files based on their date of their earliest content

58

series

contain closed files

1128

redacted:

note of exempt, exempted and restricted

accessed from 13 Oct 2020 by

buburt

who is the keeper of public records in the practice of state archives.

The Real Face of

White Australia

THE REAL FACE OF

WHITE AUSTRALIA

GET STARTED!
do I have the technical skills?

APIs are cool, but...
with the Trove API you can...
visualise newspaper searches

Visualise Trove newspaper searches over time

You know the feeling. You enter a query into Trove’s digitised newspapers search box and...

Digitised newspapers and more

Showing: 1 - 20 of at least 3,308,203 Refine search

Hmmm, 3 million results, how do you make sense of that?...

Trove tries to be as helpful as possible by ordering your results by relevance. This is great if you aim is to find a few interesting articles. But how can you get a sense of the complete results set? How can you see everything? Trove’s web interface only shows you the first 2,000 articles matching your search. But by getting data directly from the Trove API we can go bigger.

This notebook helps you zoom out and explore how the number of newspaper articles in your results varies over time by using the decade and year facets. We’ll then combine this approach with other search facets to see how we can slice a set of results up in different ways to investigate historical changes.

1. Setting things up
2. Find the number of articles per year using facets
3. How many articles in total were published each year?
4. Charting our search results as a proportion of total articles
5. Comparing multiple search terms over time
6. Comparing a search term across different states
7. Comparing a search term across different newspapers
8. Chart changes in illustration types over time
9. But what are we searching?
10. Next steps
11. Related resources
12. Further reading

If you’re interested in exploring the possibilities examined in this notebook, but are feeling a bit intimidated by the code, skip to the Related resources section for some alternative starting points. But once you’ve got a bit of confidence, please come back here to learn more about how it all works!
get your newspaper articles in bulk

Enter your search query
Use the Trove web interface to construct your search. Remember that the harvester will get all of the matched results, not just the first 2,000 you see in the web interface. Once you’re happy with your search, just copy the url and paste it below.

Query url: Enter the URL of your search

Set harvest options
By default the harvester only saves the metadata (date, page, title, newspaper etc) from the search results. If you want to save the full text content of each article, just check the Text box. You can also save PDF copies of every article by checking the PDF option, but be warned that this will slow down your harvest and generate large download files. If you want to save PDFs from large harvests, you’re probably better off installing and running the harvester on your own computer.

- Save full text
- Save PDFs (this can be slow)

Start harvest

Once your harvest is complete a link will appear to download the results as a single, zipped file. See this notebook for more information about the contents and format of the results folder.

You can also start to explore your results using this notebook.

Created by Tim Sherratt (@wragge) as part of the GLAM Workbench project.

If you think this project is worthwhile you can support it on Patreon.
Welcome to the wonderful world of GLAM data!

Here you’ll find a collection of tools and examples to help you work with data from galleries, libraries, archives, and museums (the GLAM sector), focusing on Australia and New Zealand.

What is GLAM data?

When we talk about GLAM data we’re usually referring to the collections held by cultural institutions – books, manuscripts, photographs, objects, and much more. We’re used to exploring these collections through online search interfaces or finding aids, but sometimes we want to do more – instead of a list of search results on a web page, we want access to the underlying collection data for analysis, enrichment, or visualisation. We want collections as data.

This GLAM Workbench shows you how to create your own research datasets from a variety of GLAM collections. In some cases cultural institutions provide direct access to collection data through APIs (Application Programming Interfaces) or data downloads. In other cases we have to find ways of extracting data from web interfaces – a process known as screen-scraping. Here you’ll find examples of all these approaches, as well as links to a number of pre-harvested datasets.
what does it really mean? seeing differently....
the WWI effect

https://glam-workbench.github.io/trove-newspapers/
the Hansard black hole

Comparison of Senate sitting days with number of files harvested from ParlInfo

- Yellow: Official number of sitting days
- Red: Empty Hansard files
- Blue: Harvested Hansard files

number of items described in NAA by top-level function

https://doi.org/10.5281/zenodo.3544989
number of items digitised in NAA by top-level function

https://doi.org/10.5281/zenodo.3544989
access is constructed
• how does access change?
• what can I see differently?
• what is possible that wasn't before?
• where can I go next?
I'm a historian and hacker who researches the possibilities and politics of digital cultural collections. You can follow what I'm up through my updates or on my blog.

One day a week I'm Associate Professor of Digital Heritage at the University of Canberra. The rest of the time I'm available to work on your projects – so feel free to get in touch!

My main project at the moment is the GLAM Workbench, which brings together many examples, tools, code and tutorials to help people explore the digital collections of galleries, libraries, archives, and museums.

You can find some of my other work in my open research notebook and digital heritage handbook.

If you like what I do you can support me on Patreon.