

Digital tools and such like

<https://bit.ly/digital-tools-australian-historians>

A strongly opinionated list for Australian historians by [Tim Sherratt](#)

This is not intended to be a comprehensive list. These are things that I've made or use regularly in research or teaching. Your mileage may vary.

If you have any questions head over to [OzGLAM Help](#).

Created: 15 June 2018 (last updated October 2021)

- [GLAM Workbench](#)
- [Quick hacks](#)
- [Creating and using web archives](#)
- [Trove](#)
 - [Apps and bots](#)
 - [Trove newspapers in context](#)
 - [Getting data from Trove](#)
 - [Trove data hacks](#)
 - [Trove data sets](#)
 - [Fun with Trove](#)
- [Beyond the search box](#)
- [Harvesting data](#)
 - [Other sources of data](#)
 - [Detailed lists](#)
- [Selected highlights](#)
- [Text as data](#)
- [Working with data](#)
- [Things to make](#)
- [Managing and sharing research](#)
- [Publishing collections](#)
- [Crowdsourcing](#)
- [Giving presentations](#)
- [Making websites](#)
- [Other useful tools](#)
- [Making maps](#)
- [Hosting your apps](#)
- [Must have Python packages](#)
- [Learning more](#)
- [More lists of tools and resources](#)

GLAM Workbench

The [GLAM workbench](#) includes tools, tutorials, and data to help you use and explore online collections from galleries, libraries, archives, and museums. It uses Jupyter notebooks to combine live code with text and visualisations. You can harvest and analyse data within your web browser!

The workbench currently includes examples from:

- [Trove](#)
- [National Archives of Australia](#)
- [National Museum of Australia](#)
- [NSW State Archives](#)
- [Museums Victoria](#)
- [National Archives of New Zealand](#)
- [DigitalNZ](#)
- [Te Papa](#)
- [Web archives](#) from the UK Web Archive, the National Library of Australia, the National Library of New Zealand, and the Internet Archive.
- And much more...

A number of specific GLAM Workbench tools are listed in the sections below.

Quick hacks

These userscripts change the way selected websites look and function within your browser. To use them, you first need to install a userscript manager like [Tampermonkey](#):

- [Trove Newspapers Keyboard Navigation](#) – lets you browse through digitised newspapers page by page using nothing but your arrow keys!
- [Recordsearch Show Pages](#) – shows the number of pages in digitised files in your search results from the National Archives of Australia; also stops the digitised file viewer opening in a new window, and puts a link within the digitised file viewer back to the item record.

Some more solutions to familiar problems:

- [Create persistent links in Recordsearch](#) – share or save links to items within Recordsearch without the annoying ‘session timed out’ error; note that Zotero will also create persistent links for you
 - [Save a Trove newspaper article as an image](#) (GLAM Workbench) – download a JPEG image of a complete newspaper article at a variety of resolutions; if the article stretches over multiple pages, you get multiple images!
-

Creating and using web archives

If you’re going to cite web pages, you probably want to make sure that a copy of the page is preserved somewhere. These tools can help with that by saving the page to a web archive:

- [Wayback Machine](#) – use the ‘save page now’ box to archive a web page instantly
- Wayback Machine browser add-ons are available for [Firefox](#) and [Chrome](#) – they redirect to archived versions if web pages are missing, and give you a one-click option to archive the current page
- [Archive.is](#) – another option for creating archived versions of web pages
- [Perma.cc](#) – another tool to create persistent copies of web pages; makes nicer urls, but limits apply to use.

For some projects you might want to assemble your own web archives:

- [Archive-It](#) – a subscription-based web archiving service from the Internet Archive, more for institutions than individuals
- [Conifer](#) – a web service that helps you create archives of complex interactive websites and play back the results!
- [Webrecorder](#) – a set of open source tools for the creation and playback of web archives (used by Conifer)
- [Wget](#) – a command line tool that might already be installed on your Linux or Mac machine, great for creating a working, static mirror of a website; there are many options, [see here for an example](#)

How do you make use of existing web archives? The GLAM Workbench has a [whole section](#) full of useful tools and examples, including:

- [Harvesting collections of text from archived web pages](#)
- [Create and compare full page screenshots from archived web pages](#)
- [Display changes in the text of an archived web page over time](#)
- [Find when a piece of text appears in an archived web page](#)

Trove

Apps and bots

Here are some ways of using and exploring [Trove](#) beyond the main web interface:

- [QueryPic](#) – visualise searches in Trove’s digitised newspapers (currently undergoing redevelopment – you can create new charts, but not save them online)
- [Trove places](#) – find Trove newspaper titles by place
- [Trove’s digitised journals](#) – a web app to help you search all the new journals being digitised by Trove
- [Tung Wah Newspaper Index](#) – an English-language index of two Chinese-language newspapers – the *Tung Wah News* and the *Tung Wah Times*, with links to articles in Trove
- [@TroveNewsBot](#) – explore Trove newspapers without leaving Twitter ([full documentation](#) for querying the bot)
- [Trove Twitter bots](#) – a list of Twitter bots sharing content from Trove
- [Eyes on the past](#) – eyes from Trove newspapers!
- [Culture Collage](#) (Zenlan) – explore images from multiple collections including Trove & DigitalNZ
- [To be continued... – The Australian Newspaper Fiction Database](#) (Kath Bode and Carol Hetherington, ANU) – explore serialised fiction in Trove
- [Drifter](#) (Mitchell Whitelaw) – embeds Trove newspaper article within a large-scale visualisation of the Murrumbidgee River system

Trove newspapers in context

The GLAM Workbench provides many examples of extracting and analysing data from Trove [newspapers](#), [journals](#), [books](#), [images](#), [sound](#), [lists](#), and [maps](#). Here are some examples that explore aspects of the complete collection of digitised newspapers. These all provide important context for anyone searching the digitised newspapers, helping you to understand what’s there, what’s findable, and what’s missing. The links here go to rendered versions of the original notebooks containing interactive visualisations and links to data sets.

- [Visualise the total number of newspaper articles in Trove by year and state](#) – see all of Trove’s newspaper articles at once and explore things like the impact of copyright and the WWI effect

- [Beyond the copyright cliff of death](#) – find newspapers in Trove with articles published after 1954; [here's the results](#) as a CSV file
- [Finding non-English newspapers in Trove](#) – explores ways of finding non-English language newspapers; if you're not interested in the 'how', [here's the list](#) generated by the notebook
- [Analyse rates of OCR correction](#) – explores where OCR correction has been focused within the digitised newspapers

Getting data from Trove

Trove makes a lot of its data available through an API, but not all. Here are some examples of using the API (and some screen scraping where necessary) to create useful datasets for further analysis:

- [Trove API console](#) – learn about the Trove API through lots of different sample queries that you can modify and share – all without an API key!
- [Trove Newspaper and Gazette Harvester](#) (GLAM Workbench) – download newspaper articles in bulk, including metadata, OCRd text, PDFs, and images!
- [Harvesting Parliamentary press releases via Trove Australia](#) (GLAM Workbench) – harvest metadata and full text from a search of press releases and speeches made available through Trove by the Parliamentary Library
- [Get OCRd text from a digitised journal in Trove](#) (GLAM Workbench) – download the OCRd text for every issue of a digitised journal
- [Get covers \(or any other pages\) from a digitised journal in Trove](#) (GLAM Workbench) – download selected pages as images from all issues of a digitised journal

Trove data hacks

- [Getting random items from Trove](#) (GLAM Workbench) – a variety of hacky workarounds to retrieve a random-ish item from Trove
- [Trove proxy](#) – provides handy shortcuts to programmatically retrieve the PDF of a digitised newspaper article and find random things

Trove data sets

Here are some selected, pre-harvested, ready-to-go datasets, you'll find more in the GLAM Workbench:

- [Geolocated Trove newspaper titles](#) – CSV file of newspapers in Trove with spatial coordinates of places of publication/distribution (used in the Trove places app)
- [ABC Radio National programs](#) (GLAM Workbench) – CSV files containing program metadata from ABC Radio National harvested through the Trove

- [OCRd text from Trove's digitised journals](#) (GLAM Workbench) – text from 33,035 issues of 719 journals
- [OCRd text from Trove's digitised books](#) (GLAM Workbench) – text from 9,738 books (GLAM Workbench)
- [Editorial cartoons from *The Bulletin*, 1886-1952](#) (GLAM Workbench) – 3,471 images available for download as jpgs, or compiled into PDFs for easy browsing
- [Australian Women's Weekly front covers, 1933 to 1982](#) (GLAM Workbench) –
- [Parliamentary press releases relating to refugees](#) (GLAM Workbench) – CSV formatted metadata for the collection and individual text files with the content of each document
- [To be continued... – The Australian Newspaper Fiction Database](#) (Kath Bode and Carol Hetherington, ANU) – lets you download metadata and text from their database of serialised fiction

Fun with Trove

- [Headline Roulette](#) – how well do you know your Australian history?
- [The Vintage Face Depot](#) – find a new face from our extensive range of vintage visages
- [Create 'scissors and paste' messages from Trove newspaper articles](#) (GLAM Workbench) – snips printed words out of newspaper articles and assembles them into a downloadable image

Beyond the search box

Some alternative/unofficial/experimental ways of exploring Australian cultural heritage collections. See also [More GLAM tools and interfaces](#) in the GLAM Workbench.

- [Colonial Frontier Massacres in Australia, 1788-1930](#) (University of Newcastle)
- [Historic Hansard](#) – Commonwealth of Australia parliamentary debates 1901-1980 presented in an easy-to-read format
- [Real Words :: Imagined Tweets](#) – Historical Parliamentary interjections re-imagined as tweets
- [GLAM CSV Explorer](#) (GLAM Workbench)– visualise the contents of hundreds of data files shared by Australian GLAM organisations
- [Prints and Printmaking](#) (Ben Ennis Butler) – a series of innovative visualisations of prints in the National Gallery of Australia
- [Discover the Queenslander](#) (Mitchell Whitelaw and Geoff Hinchcliffe) – covers of *The Queenslander*
- [Loom](#) (DXLab, SLNSW) – photo and subject browser
- [The Real Face of White Australia](#) – faces from White Australia records in NAA

- [Closed access](#) – explore the NAA files you're not allowed to see
 - [Redacted](#) – explore redactions in ASIO files from NAA
-

Harvesting data

An assortment of harvesting tools that turn websites into datasets.

- [RecordSearch harvesting](#) – harvest records and images from a search in the National Archives of Australia (GLAM Workbench)
 - [Archway harvesting](#) (GLAM Workbench) – get item-level data out of Archives New Zealand's online database
 - [Papers Past](#) (GLAM Workbench) – harvest New Zealand newspapers
 - [Twarc](#) – a command-line tool for saving collections of Tweets
 - See also [Getting data from Trove](#)
-

Other sources of data

Detailed lists

For GLAM and HASS related data sources, including downloadable datasets and APIs, see the lists below:

- [Sources of Australian GLAM data](#) (GLAM Workbench)
- [GLAM datasets from government data portals](#) (GLAM Workbench) – details of hundreds of GLAM-related datasets from libraries, archives and museums around Australia
- See also [Trove datasets](#)

Selected highlights

- [Commonwealth Hansard, 1901-1980](#) – repository of XML files download from the Parliament of Australia website
- [Western Australia Character Books, 1850-57](#) – CSV file with records of 358 convicts who were tried in London
- [Papers Past Newspaper Open Data Pilot](#) – downloadable NZ newspaper datasets organised by title and year

- [State Archives of NSW indexes](#) (GLAM Workbench) – CSV formatted versions of 64 indexes with 1,499,259 rows of data
 - [Tribune negatives collection](#) – CSV file of metadata from SLNSW catalogue describing 60,000 images
 - [NAA series relating to the White Australia Policy](#) (GLAM Workbench) – summaries and complete metadata (in CSV format) for 23 series
 - [PM Transcripts repository](#) (GLAM Workbench) – index in CSV format and XML transcripts of more than 20,000 speeches, media releases, and interviews by Australian Prime Ministers provided by the Department of the Prime Minister and Cabinet
 - [Election Speeches](#) (Museum of Australian Democracy) – downloadable zip file with complete speeches in text format
-

Text as data

Some web-based tools for starting to explore texts as data.

- [Election Speeches](#) (Museum of Australian Democracy) – a great example to start thinking about word frequency analysis
 - [Word Counter](#) – a simple tool to explore word/ngram frequencies, part of Data Basic
 - [Voyant Tools](#) – amazingly powerful web platform for text analysis
 - [Google Ngram viewer](#) – need to be aware of its limitations, but a good tool to play around with in class (particularly when you get them searching for ‘fuck’)
 - [Bookworm](#) – ngram exploration across a variety of text collections
 - [Overview](#) – look for patterns and connections in large collections of documents
-

Working with data

Tools to clean and visualise your data.

- [Data Basic](#) – a collection of simple online tools that get you thinking about the nature of data and what you can do with it
- [Open Refine](#) – the swiss army knife of data cleaning and linking; if you’re working with messy data, you *need* Open Refine
- [Jupyter](#) – explore data in your browser, combining code, text, visualisations and more

- [Pandas cookbook](#) – Jupyter and Pandas are my new go-to combination for any sort of data exploration, this cookbook provides a good introduction to the possibilities of Pandas
 - [Plot.ly](#) – I’ve used Plot.ly a lot for charts and visualisations, but they’ve been shifting away from research into the corporate market and no longer seem to have a free account (although their software is still open source)
 - [RAWgraphs](#) – the easiest way to go from CSV to data visualisation, just drop in your data and choose a visualisation
 - [The Vega-Lite family](#) – my new favourite framework for charts and visualisations, not as quick to get started as Plot.ly, but very flexible
 - [nteract](#) – a desktop app for building Jupyter style notebooks for data exploration
-

Things to make

Making stuff is one of the best ways to learn.

- [Trove Twitter bots for all!](#) – create and customise your very own Trove-powered Twitter bot
 - DIY Headline Roulette – remixable versions available on [GitHub](#) and [Glitch](#)
 - [DIY Trove exhibition](#) – with nothing more than a Trove list and a GitHub account you can create your own online exhibition in minutes!
 - [Cheap bots done quick!](#) – easy and fun way to make your own Twitter bot
 - [Glitch](#) – find something interesting, then remix it!
 - [Twine](#) – create and share your own non-linear narratives and text-based games
-

Managing and sharing research

Some tools to manage your research:

- [Zotero](#) – much more than a reference manager, use it to manage and organise your research; Zotero comes with built-in translators for Trove newspapers and NAA RecordSearch that automatically capture metadata and PDFs
- [Tropy](#) – describe and organise all those photos you take in archives
- [Hypothes.is](#) – annotate web pages and PDFs; [create direct links](#) to text fragments!

Some options for sharing your research outputs:

- [Open access publishing for Australian historians](#) – another google doc collecting information and resources to help you make your research open access
 - [Figshare](#) & [Zenodo](#) – open, all-purpose repositories for sharing your research outputs, including presentations, data and code
 - [Humanities Commons](#) – like Academia.edu but run by scholars who won't sell your data
-

Publishing collections

- [Omeka and Omeka-S](#) – publishing platform for online collections and exhibitions, Omeka S is the next-gen version with many new features
 - [Omeka.net](#) – hosted version of Omeka
 - [Wakelet](#) – capture and publish collections of Tweets and other online resources
-

Crowdsourcing

Build your own crowdsourcing application:

- [Zooniverse project builder](#) – upload your data, define your tasks, and mobilise your volunteers
- [Scribe API](#) – a framework for crowdsourced transcription of historical documents created by Zooniverse and NYPL (see my [updated version](#) of the code)
- [Scripto](#) – a transcription plug-in for Omeka

Some crowdsourcing projects you can contribute to:

- [The Real Face of White Australia](#) – see Scribe in action and help us transcribe records!
- [Criminal Characters](#)
- [PROV Map Warper](#)

Giving presentations

I like to live dangerously and embed live sites, charts, and visualisations in presentations. Reveal.js (in its various forms) makes it easy to embed iframes. No more cheesy animations, just real life web apps.

- [Reveal.js](#) – an excellent Javascript library for building presentations
- [Slides.com](#) – hosted version of Reveal.js that makes it very, very easy to create slideshows that can be run and shared online
- [RISE](#) – Jupyter notebooks + Reveal.js = live code in presentations!

Making websites

- [GitHub pages](#) – create simple websites, hosted for free on GitHub
- [Jekyll](#) – a static site generator, use it with GitHub pages or on your own site
- [Ed. a Jekyll theme for minimal editions](#) – create scholarly editions of works using Jekyll
- [Reclaim Hosting](#) – if you need a good, cheap, reliable web host; one click installs for Omeka and Omeka-S!

Other useful tools

- [JuxtaposeJS](#) – online tool for creating side-by-side comparisons of images or maps
- [TimelineJS](#) – online tool for creating nice-looking timelines that you can embed in your own site

Making maps

- [Carto](#) – used to be great for combining data and maps, but they've taken a turn to the corporate side and limited the free options
- [MapBox](#) – powerful platform for creating and styling your own custom maps
- [Leaflet](#) – the go-to library for making maps with Javascript
- [StoryMapJS](#) – tell stories on maps
- [Neatline](#) – an Omeka plug-in that enables you to combine maps, text, and documents in interesting and creative ways

Hosting your apps

- [Heroku](#) – there's a generous free tier that's enough for experimental apps, then scale up to paid plans as required
- [Glitch](#) – create, remix, share, deploy, all in the one free platform
- [Reclaim Cloud](#) – a new-ish service from Reclaim Hosting lets you host more complex applications
- [Binder](#) – run your Jupyter notebooks in the cloud!

Must have Python packages

These are the packages I find myself installing most frequently.

- [Pandas](#) – do magic with tabular data
- [Altair](#) – great looking charts with a minimum of code
- [BeautifulSoup](#) – turn HTML pages into things you can manipulate with code
- [Requests](#) – easiest way to make HTTP requests
- [Flask](#) – build web apps quickly and painlessly
- [TextBlob](#) – a text analysis package that is much easier to understand than NLTK
- [Jupyter](#) – use all of the above and more in your browser, then share it all with others

Learning more

- [The Programming Historian](#) – easy-to-follow tutorials on a wide variety of topics aimed directly at digital historians (both beginners and advanced)
- [Library Carpentry](#) – aimed at librarians but plenty of useful information on digital tools and skills for historians as well
- [Trove tips & tricks](#) – slides from an old workshop (somewhat out-of-date after the Trove development)
- [Trove as a platform for digital research & creativity](#) – more slides, these ones move beyond the web interface to explore the API
- [Digital Heritage Handbook](#) – my collection of workshops, tutorials, documentation etc (needs a refresh, but still some useful bits and pieces)
- [The Open Digital Archaeology Textbook](#) – includes many useful Jupyter notebooks

More lists of tools and resources

- [Miriam Posner's Highly Opinionated Resource Guide](#)
- [TAPOR 3](#) – now incorporating the old DiRT Directory of DH tools