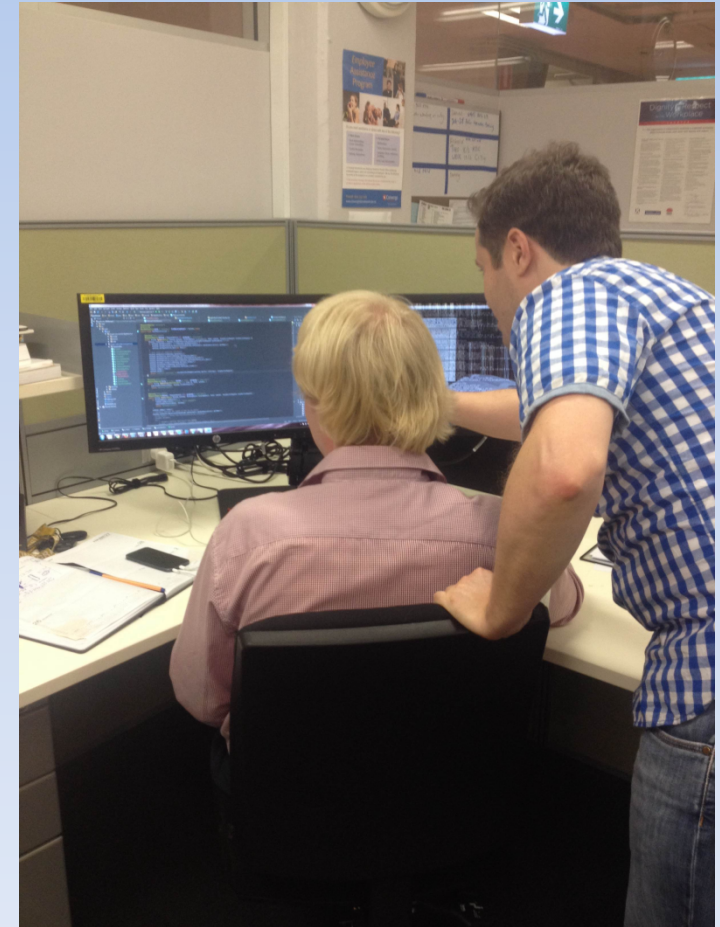# State Records
Digital Archives

Digital State Archive
OpenGov NSW
Data NSW

- 2011: we started the project to build infrastructure and processes that would enable us to accept, preserve and make available digital Government records as State archives

- Mid-2014: we move from 'project phase' to 'business as usual' status

What are we archiving?

- Any digital information that is kept by a government entity as evidence of official business, that:
  - is 'required as State archives' in an authorised retention and disposal authority.
  - is no longer in active use
  - has access rules in place
- Could be in the form of:
  - documents, spreadsheets, PPTs;
  - case management systems;
  - email
  - scanned copies of paper records or digitised A/V material
  - EDRMS records

What we are not about

- We are *not* about digitisation (conversion of paper to digital images) – although some agencies who do this may send the digital copies to us eventually

- Our value is *not* in storage or back up. These things can be done by lots of providers.

- We are about analysing, migrating, preserving, contextualising and providing access to digital records in digital recordkeeping systems

Pilot projects

- Records of former Premier Kristina Keneally
- Special Commission of Inquiry into Electricity Transactions database
- Business Registration System / Business Registration Imaging System
- Records of former Premier Nathan Rees
- Records of former Premier Bob Carr
- Farrer Memorial Oration
- Public Works Oral History recordings
- Thredbo Landslip Coronial Inquiry database

The Challenges of Digital Preservation

Vint Cerf

"I worry a great deal about that... **Old formats of documents** that we've created or presentations may not be readable by the latest version of the software because **backwards compatibility is not always guaranteed**.

"And so what can happen over time is that even if we accumulate **vast archives of digital content**, we may not actually know what it is."

http://www.bbc.com/news/science-environment-31450389 (13 Feb 15)

Example: Special Commission of Inquiry into Electricity Transactions

- 40,813 file objects (approx 35 Gb):

  - .ZIP/JAR File, 1
  - Adobe Acrobat (PDF), 28,920
  - Compuserve GIF, 235
  - Generic Password Protected Microsoft Office 2007 Document, 13
  - Internet HTML, 14
  - Internet Mail Message, 6
  - JPEG File Interchange, 12
  - Microsoft Access 7, 2
  - Microsoft Excel 2000, 901
  - Microsoft Excel 2002, 2
  - Microsoft Excel 2007/2008, 396
  - Microsoft Excel 2007/2008 Macro Enabled Workbook, 16
  - Microsoft Excel 5.0/7.0, 6
  - Microsoft Excel 97/98/2004, 1
  - Microsoft PowerPoint 2007/2008, 85
  - Microsoft PowerPoint 97-2004, 77
  - Microsoft Project 2000/2002/2003, 3

  - Microsoft Word 2000, 85
  - Microsoft Word 2002, 82
  - Microsoft Word 2003/2004, 1,570
  - Microsoft Word 2007/2008, 281
  - Microsoft Word 2007/2008 Macro Enabled Document, 1
  - Microsoft Word 97/98, 4
  - MS Outlook, 7,689
  - Password Protected Microsoft Excel 2007/2008, 3
  - Portable Network Graphics Format, 1
  - Progressive JPEG, 3
  - Tagged Image File Format, 39
  - Text Mail, 2
  - Unknown format, 335
  - vCard, 23
  - Windows Bitmap, 4

- Relativity database metadata export in XML (100 Mb)

The challenges we encountered

- Diversity: huge variety of systems, formats and dependencies in use in the NSW Government

- Quality: some very well managed systems, lots of documentation, some very poorly managed

- Interconnected systems, not individual objects

- Integration: the need for us to manage and search across all these diverse systems

- Support: commitment to undertake the work necessary to preserve and migrate

Evgeny Morozov

"Imperfection, ambiguity, opacity, disorder, and the opportunity to err, to sin, to do the wrong thing: all of these are constitutive of human freedom, and any concentrated attempt to root them out will root out that freedom as well"

To Save Everything Click Here: The Folly of Technological Solutionism

Migration Methodology

NSW State Records Digital Archives

The Digital Archives Migration Methodology supports the transfer of digital records from NSW Government agencies to the Digital State Archive. Rather than adopting a single approach for all such transfers, State Records NSW defines custom migration plans to suit the particular requirements of different sets of records. Each transfer is managed as a separate project. The methodology is a framework to guide these projects.

By blending project management and data migration techniques, the Digital Archives Migration Methodology provides a structured and planned approach to each migration project. It also permits flexibility, for dealing with many types of migrations, from very simple ones to complex ones involving many record types and stakeholders.

1. Project Planning Phase

The Project Planning phase establishes a framework for the migration project.

The purpose of this phase is to define the project goals and identify stakeholders, risks, and resources. The depth of planning required in this phase will vary depending on the complexity of the project. For example, a project involving the transfer of a single audio file might be very small in comparison with a project involving the transfer of a business or email system.

The key deliverable of this phase is the project plan.

Tools

Basecamp *Adopted*
https://basecamp.com
Basecamp is a web-based project management tool. It is used to collaborate with project participants, plan and schedule actions, and share documentation.

Digital Archives Migration Methodology *Built*

Key Relationships

State Records Act 1998
S. 29 of the State Records Act allows State Records to issue guidelines about how records are to be made available to it. The guidelines apply to records in any format, including digital records.

2. Migration Planning Phase

The goal of the Migration Planning phase is to develop the migration plan.

The migration plan is a document that identifies and documents the activities to be carried out during the migration of a recordkeeping system into the Digital State Archive. Those activities include file format migration, metadata mapping, and data transformation.

Like project plans, migration plans are tailored to suit the requirements of the particular project. But this doesn't mean re-inventing the wheel each time: the decisions and lessons learned in each project are documented and can be re-used in subsequent projects.

Tools

DROID/PRONOM *Adopted*
http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx
The National Archives UK's technical registry and file format identification tool are used for canonical identification of file formats.

Apache Tika *Adopted*
http://tika.apache.org/

ExifTool *Adopted*
http://www.sno.phy.queensu.ca/~phil/exiftool/
Metadata extraction tools such as Tika and ExifTool are used to supplement agency-provided metadata.

Preservation Pathways Registry *Built*
http://www.records.nsw.gov.au/digitalarchives/pathways/
When digital records are migrated to the digital archives, file formats are assessed for their longevity and accessibility. In some cases a transformation is recommended. Transformations are registered in Preservation Pathways with information about the input and target file formats (using IDs from the National Archives UK's PRONOM registry as well as information about the tool or process used to perform the transformation). If file format IDs are not available from the PRONOM registry, then temporary State Records NSW IDs are registered pending the creation of a PRONOM ID.

The Preservation Pathways Registry is a Java web application.

Metadata Registry *Built*
http://www.records.nsw.gov.au/digitalarchives/metadata/
The Digital Archives Metadata Registry allows Digital Archives staff to progressively register preferences for published metadata terms (e.g. Dublin Core) to represent common metadata elements in the digital archives. It also allows Digital Archives staff to progressively coin new terms (by providing a URI and description) to represent metadata elements in the digital archives for which no suitable published term can be identified.

The Metadata Registry is implemented with JSON Schema and using Github.

Key Relationships

Disposal authorities
Part 3 of the State Records Act 1998 prohibits the disposal of State records except where it is authorised. Under the Act, State Records can give permission for disposal. The usual means by which State Records permits disposal is through the approval of retention and disposal authorities.

Only digital records required as State archives under an authorised disposal authority may be transferred to the Digital State Archive.

Access Directions
Part 6 of the State Records Act creates a framework for regulating public access to State records which have been in existence for at least 30 years (the 'open access period'). Public offices are required to make an access direction (to determine whether the records are open or closed to public access) for all their records which are in the open access period.

Agencies transferring records to the Digital State Archive must ensure that those records are covered by current access directions.

Series control system
State Records implements the Australian Series System to describe and control the State Archives. Digital archives are linked to this system by links to agencies and series.

3. Migration Phase

During Migration phase, the migration plan is executed. It is in this stage that any necessary preservation activities are performed.

Tools

Aspose *Adopted*
http://www.aspose.com/
A set of file format APIs for Java. Used for transformation of a number of common file formats.

EMC Isilon *Adopted*
http://www.emc.com/domains/isilon/index.htm
EMC Isilon scale-out network-attached storage is used to store digital archives. Advantages of this system include scalability, integrity and automated replication.

Pairtree *Adopted*
https://wiki.ucop.edu/display/Curation/PairTree
A modified version of California Digital Library's Pairtrees for Object Storage protocol is used to manage the storage of digital objects. Pairtree is a filesystem hierarchy.

MongoDB *Adopted*
http://www.mongodb.org/
Metadata is stored as JSON documents in the filesystem and copied into a MongoDB instance for search and reporting

Apache Solr *Adopted*
http://lucene.apache.org/solr/
Full-text is extracted wherever possible. This full-text is stored in a Solr instance for search.

Migrate tool *Built*
A workflow tool that assigns unique identifiers to digital objects, validates consignments, and moves records into the Digital State Archive.

The Migrate tool is a Java command-line application.

Key Relationships

Search
http://search.records.nsw.gov.au/search
The Digital Archives team at State Records is responsible for Search, the main finding aid for the State Archives collection. Contents of the Digital State Archive are discoverable via Search.

4. Project Closure Phase

The Project Closure phase closes the project and identifies any required post-project activities (such as the disposal of source records).

Tools

Basecamp *Adopted*
http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx
Basecamp is a web-based project management tool. It is used to collaborate with project participants, plan and schedule actions, and share documentation.

Receipt *Built*
http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx
This tool provides agencies with a human and machine readable manifest of digital records successfully migrated into the Digital State Archive.

The Receipt tool is a Java command-line application.

Key Relationships

GA33 Source Records that have been migrated
This is a disposal authority that provides for the authorised disposal of State Records that have been used as the source records for successful migration projects. It is under this authority that public offices are permitted to destroy source records post-transfer to the Digital State Archive. This authority mandates a minimum six months retention period for quality assurance purposes.

Digital archives migration methodology

![State Records Digital Archives NSW Government]
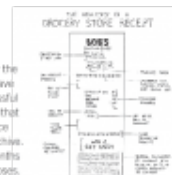
Phase 1 Project Planning	Phase 2 Migration Planning	Phase 3 Migration	Phase 4 Project Closure

- Who does what
- When
- Outcomes

- Analyse and prepare the records for migration

- Carry out the migration
- Transfer of control

- Confirm migration success
- Deal with source records
- Access arrangements

Methodology flexible enough to cope with simple or complex; custody or non-custody; hybrid or all digital

Project planning

- What is the scope of this project?

- Who should be involved?

- What will it cost? What resources are necessary?

- How long will it take?

- What value will the agency realize?

Migration planning

- What are we going to do with this format?
- What metadata resides in this system? What metadata do we keep? How does it relate to the metadata in other series in the DSA?
- How will we maintain the relationships between interlinked record systems?
- What will researchers of the future want to know about this system?
- What queries will the responsible agency want to run?
- What was the system's interface / look and feel?
- What documentation is there of this system?
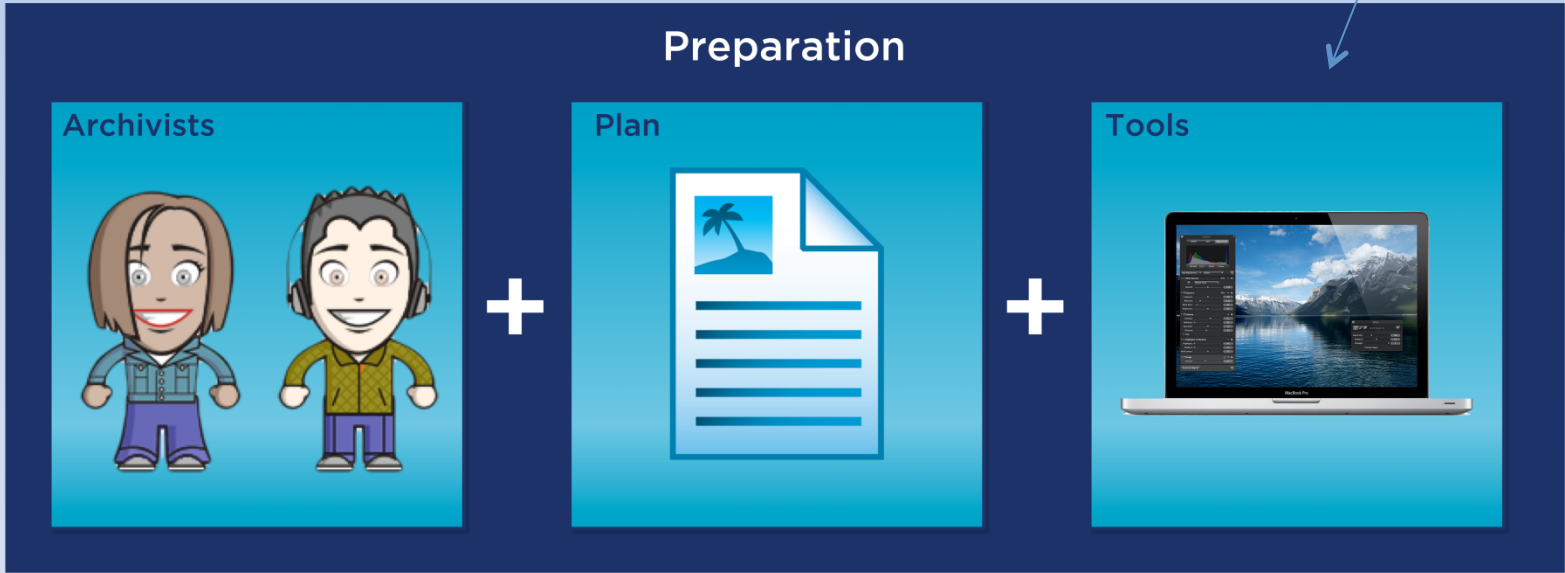
File format identification

Tagged Image File Format
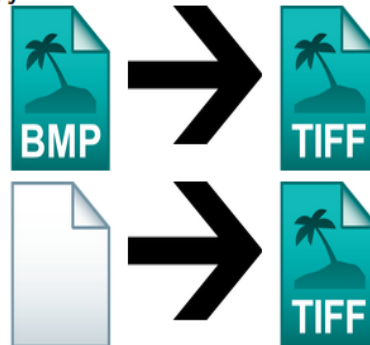

Name:

   Tagged Image File Format

Description:

The Tagged Image File Format (TIFF) is a raster image format originally developed by the Aldus Corporation, primarily for u... scanning and desktop publishing. When Adobe Systems Incorporated purchased Aldus in 1994, they acquired the rights to th... format and have maintained it since then. TIFF files comprise three sections: an Image File Header (IFH), an Image File Direct... (IFD), and the image data. TIFF files can contain multiple images (multi-page TIFF), and each image has a separate IFD. The... always appears at the beginning of the file, and is immediately followed by a pointer to the first IFD. The IFD contains metad... which describes the associated image, stored as a series of tags. The IFD also contains a pointer to the actual image data. TI... supports colour depths from 1 bit to 24 bit (e.g. monochrome to true colour), and a wide range of compression types (RLE, L... CCITT Group 3 and Group 4, and JPEG), as well as uncompressed data.

Generated:

   true

Extension:

   tiff

infourl:

   info:srnsw/formats/1010

Pathways In:

State Records
Digital Archives

NSW GOVERNMENT

| Home | About | Contact | Visit | The State Archives Collection | Recordkeeping NSW public sector | Government |

Digital Archives

Pathways

List all Pathways

Purposes

List all Purposes

Format

List all Formats

External Format Identifier

List all External Format Identifiers

Transformer

List all Transformers

Pathway	Input Format	Transformer	Output Format
	Microsoft Word (Generic)	Optimus Prime	Acrobat PDF/X - Portable Document Format - Exchange 1:1999
	Microsoft Word for Macintosh Document	Ironhide	Microsoft Word (Generic)
	Microsoft Word for MS-DOS Document		Microsoft Word (Generic)
	OS/2 Bitmap	Steeljaw	Tagged Image File Format
	Macintosh PICT Image		Tagged Image File Format

List results per page: 5 10 15 20 25 | Page 1 of 1

Digital Archives - Metadata registry

https://github.com/srnsw/metadata

[Download Schema](#)

12 days ago

title	string
id	string
consignmentId	string
metadataId	string
created	W3CDate

W3CDate

string `//(\d{4}-[01]\d-[0-3]\dT[0-2]\d:[0-5]\d:[0-5]\d\.\d+)|(\d{4}-[01]\d-[0-3]\dT[0-2]\d:[0-5]\d:[0-5]\d)|(\d{4}-[01]\d-[0-3]\dT[0-2]\d:[0-5]\d)//`

| creators | array |
| recipients | array | agent |

agent

one of [object] [object]

| person | person |

person

| surname | string |

How does an agency transfer digital archives?

- They contact us by usual methods or records at risk are identified

- We confirm whether the records are eligible (same rules as for paper)

- Once it looks like going ahead we log them into our project management tool and start the migration project

- Projects might take a few weeks or months, depending on the complexity

What about access?

- Records that are open to public access under an Access Direction to be available via our usual Search tools

- Data extraction will mean exciting possibilities for users of the archives

- Closed records to be available to controlling agency for responding to GIPA, business use etc.

The Opportunities

OpenGov NSW

Search   **About**   **Faqs**

✉ Contact

Agency Login

3525 publications

Contains

Enter what you are looking for

☐ Just search the title

Published by

Type in an agency

Publication type

All types ▾

Published after

Any date 📅

Published before

Any date 📅

Search

Welcome to OpenGov NSW

This is where you will find information published by NSW Government agencies, including Annual Reports and open access information released under the Government Information (Public Access) Act 2009 (GIPA Act). For government datasets, you should visit **Data NSW** or the **NSW Spatial Data Catalogue**. If you would like to browse archival records, please visit the **State Records NSW** website.

Attention: NSW Caretaker Period has commenced
The caretaker period for the NSW Election commenced on 6 March 2015.
Accordingly, no ministerial press releases or related information issued by the Government from this date will be available on this website. For copies of recently-issued ministerial press releases of information on the election policies of any political party as they relate to this department/agency or its portfolio area, please go to the website of the relevant political party.

Recently Published

Popular

🏠 / Datasets

▼ Organisations

Department of Educa... (57)

Roads and Maritime ... (24)

Office of State Rev... (23)

Transport for NSW (20)

Department of Plann... (19)

Bureau of Transport... (17)

Anti-Discrimination... (15)

Office of Environme... (11)

Land and Property I... (10)

State Records Autho... (9)

Show More Organisations

▼ Groups

Infrastructure and ... (42)

Education and Training (34)

Business and Indust... (24)

⊕ **Download Summary Table**

Search datasets...   🔍

311 datasets found

Order by: Relevance ▼

BTS Data Visualisations

The following Bureau of Transport Statistics data visualisations are available: Travel Zone Explorer Journey to Work Explorer Local Government Area Profiler...

`HTML`

NSW and Sydney Transport Facts

The Transport Facts dashboard provides important top-level transport statistics at your fingertips. Categories include: Population Economy Labour Vehicles...

`HTML`

Landuse Planner

The Landuse Planner provides a detailed look at population, employment and workforce forecasts by area, and allows for the production of thematic maps.

`HTML`

▼ **Organisations**

NSW Health (3)

Office of Finance a... (2)

Trade & Investment ... (1)

Fire & Rescue NSW (1)

Bureau of Health In... (1)

▼ **Groups**

Health (2)

Conservation and En... (1)

▼ **Tags**

health (4)

government (2)

NSW (2)

Media (2)

natural-resources (1)

mental health (1)

media-releases (1)

⊕ **Download Summary Table**

media releases 🔍

8 datasets found for "media releases"

Order by: Relevance ▼

Resources & Energy Media Releases

Media Releases from Trade & Investment (Resources & Energy).

`PDF`

Media Releases from NSW Health

Machine readable feed of media releases from NSW Health.

`rss`

Fire & Rescue NSW Media Releases

RSS feed of Fire & Rescue NSW media releases.

`rss`

Bureau of Health Information Media Releases

RSS feed of latest NSW public hospitals performance report media releases from the Bureau of Health Information. For all BHI media releases, visit...

Contains

Enter what you are looking for

☐ Just search the title

Published by

Type in an agency

Publication type

All types ▾

Published after

Any date 📅

Published before

Any date 📅

Search

We found 272 matches for:

Publications published by any agency, of type **Media Release**, within any date range

Refine results

You can use the filters below to further refine the results

Published by

All Agencies ▾

Publication type

Media Release (272) ▾

Published after

Any date 📅

Published before

Any date 📅

Update results

Sort by ● **Relevance**   ○ **Earliest**   ○ **Latest**

$1.15 Billion Tunnelling Contract Awarded On North West Rail Link

Agency: Department of Premier and Cabinet

Published: 25/06/2013   **Type:** Media Release   ⬇ Download: PDF (16568

$2 Million For Emu Plains Pinch Point

Agency: Department of Premier and Cabinet

Published: 01/07/2013   **Type:** Media Release   ⬇ Download: PDF (16467

$2 Billion Investment Secured For Barangaroo

Agency: Department of Premier and Cabinet

Published: 08/07/2012   **Type:** Media Release   ⬇ Download: PDF (14598

State Records
Digital Archives

Home   Objects   **Series**   All Series

Media releases from the office of Kristina Keneally, Premier of New South Wales

This series consists of media releases from the office of Kristina Kenneally during her time as Premier of New South Wales between April 2009 and March 2011.

The media releases are arranged in chronological order and were received electronically in .PDF format.

The media releases relate to matters in the Premier's portfolio and consist of policy announcements, announcements regarding event attended by the Premier, recent successes within the government as well as new projects and initiatives.

Each press release is 1-2 pages in length covering a single topic or issue and containing a media contact name and number at the end

Title	Created	Series
NSW Unis Research And Development	2010-07-13	20,282
Rural Health	2010-03-16	20,282
South Coast Correctional Centre	2010-11-12	20,282
Royal Easter Show	2010-04-01	20,282
Barangaroo Transport	2010-12-23	20,282
LA Galaxy And Newcastle Jets	2010-10-20	20,282
Mary MacKillop Bridge Display	2010-10-11	20,282

NSW GOVERNMENT | State Records

| Home | About | Contact | Visit | The State Archives Collection | Recordkeeping NSW public sector | Government Records Repository |

You are here :: Home → Search

[press releases] **Collection search**

Filter by year

1775+ (35)

Filter by series

📄 Registered files (2)

📄 Media releases from the office of Kristina Keneally, Premier of New South Wales (1)

📄 Press releases from the office of Pam Allan, Minister for the Environment 1995-1999 (1)

📄 State Coal Mines draft leases (1)

📄 Video tape, audio tape, film and other media productions for Department of Education [VT series] (1)

Record series and items

📁 **Press releases [Department of Lands]**

...These ministerial **press releases** concern such matters as: Crown land subdivisions; drought aid; the building of new schools; assistance for primary...

📁 **Press releases [Premier's Department]**

...The series comprises signed copies of **press releases**, prepared for publication in various New South Wales newspapers, by Premiers of New South Wales...

📁 **Indexes to press releases [Department of Lands]**

...These indexes give the subject matter and page number of departmental **press releases**. (They index the **press releases** located at 19/17675-76.) (19...

📁 **Newspaper clippings and press releases [Nurses Education Board]**

...Newspaper clippings, 1975, **press releases**, 1976-78 (18/1429.3). 1 bundle. Note::
This description is extracted from Concise Guide...

📁 **Press releases from the office of Pam Allan, Minister for the Environment 1995-1999**

Functions and activities

No results

Agencies and people

👥 Community Relations and Publications Section [NSW Planning and Environment Commission]

👥 Circular Quay Engineering Committee (1938 - 1939) / Circular Quay Supervisory Committee (1939 - 1964)

👥 Media Unit [NSW Police]

👥 Stockton Hospital for the Insane (1910-1917) Stockton Mental Hospital / Stockton Hospital (1917-1989) Stockton Centre (1989-)

👥 Rabbit Island Hospital for the Insane (1910-1917) Rabbit Island Mental Hospital (1917-1936) Peat and Milson Islands Mental Hospital / Peat and Milson Islands Hospital (1936-1973) Peat Island Hospital (1973-1989) Peat Island Centre

Simon Elvery ✓
@drzax

Follow

Hi @datansw. You do a great job, thank you. However, it would be nice to correct this:

📍 South Brisbane, Brisbane

You are here: **Home** / **Search** / **Content** / Search

Search

Enter your keywords   elections   **Search**

Your search yielded no results

- Check if your spelling is correct.
- Remove quotes around phrases to search for each word individually. *bike shed* will often show more results than *"bike shed"*.
- Consider loosening your query with *OR*. *bike OR shed* will often show more results than *bike shed*.

The 2014 Victorian state election has been run and won. While there are still some recounts underway, we have enough data from the Victorian Electoral Commission to look at the count in more detail.

Explore the interactive map and toggle between the 2010 and 2014 results to see how the vote changed on a booth-by-booth level.
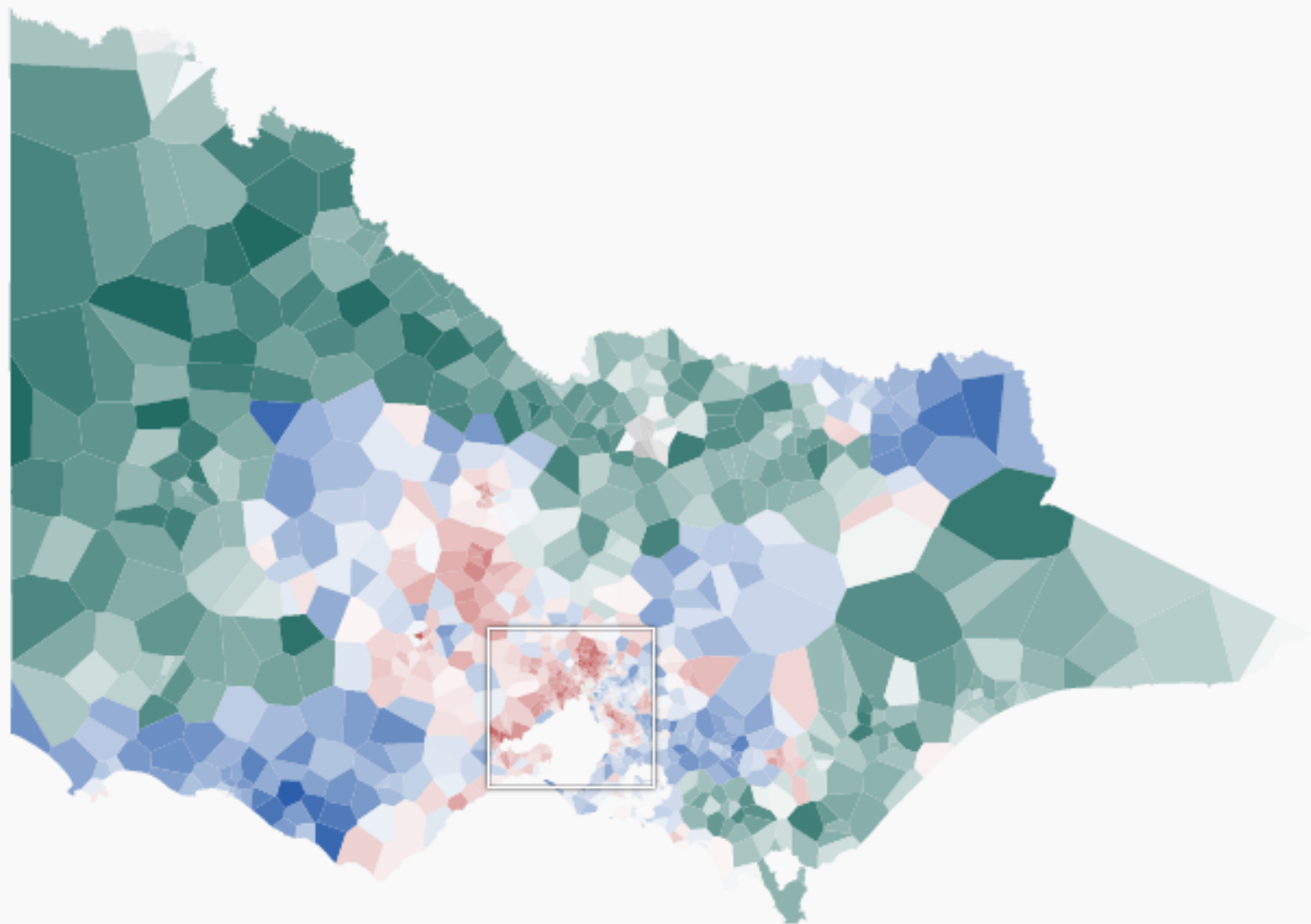
Find an Electorate   Suburb or postcode: [   ]

| 2010 | 2014 |

☐ Show only booths where the leading two-candidate preferred party changed.

STRONGER PARTY SUPPORT →

ALP ▭ LIB ▭ NAT ▭ IND ▭
GRN ▭

Links

http://www.records.nsw.gov.au/digitalarchives

http://futureproof.records.nsw.gov.au/

http://www.opengov.nsw.gov.au

http://data.nsw.gov.au