





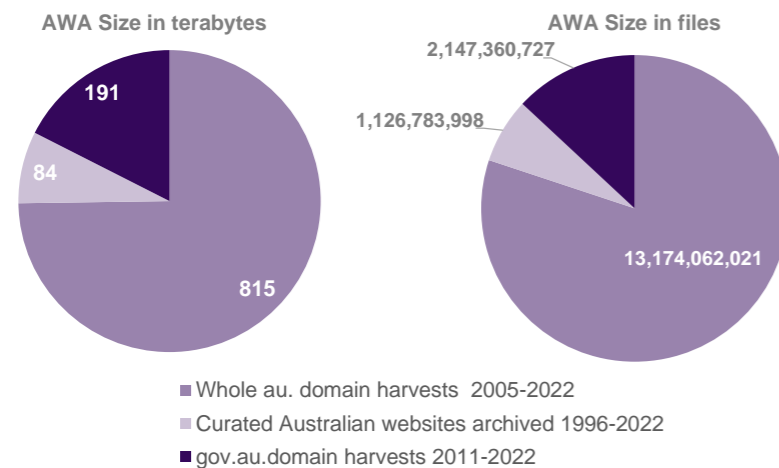


Australian Web Archiving

PAST

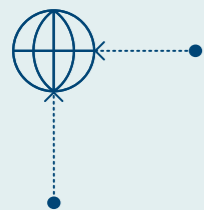
-  Selective collecting of individual websites **since 1996** by the Library in collaboration with other Australian partners.
-  Domain harvesting of the .au domain **since 2005** in collaboration with the Internet Archive.
-  Bulk harvesting of Australian Government websites undertaken in-house **since 2011**.
-  The ability to open up the Library's complete archive of .au websites to the public was restricted by both technological capabilities and legal constraints.
-  2016 expansion of the *Copyright Act 1968* legal deposit provisions to include digital material AND internal investment in our technological capabilities.
-  In 2019 the Library opened its full archive including the .au domain for public search through Trove. The archive was renamed The Australian Web Archive (AWA) at this time.



INTERNATIONAL BEST PRACTICE



The NLA continues to have a presence internationally and was one of the 12 founding members of the International Internet Preservation Consortium (IIPC) - established in 2003.



As a member of the IIPC the Library has worked to improve the tools, standards and best practices of web archiving while promoting international collaboration and the broad access and use of web archives for research and cultural heritage.

PRESENT

Collection Building

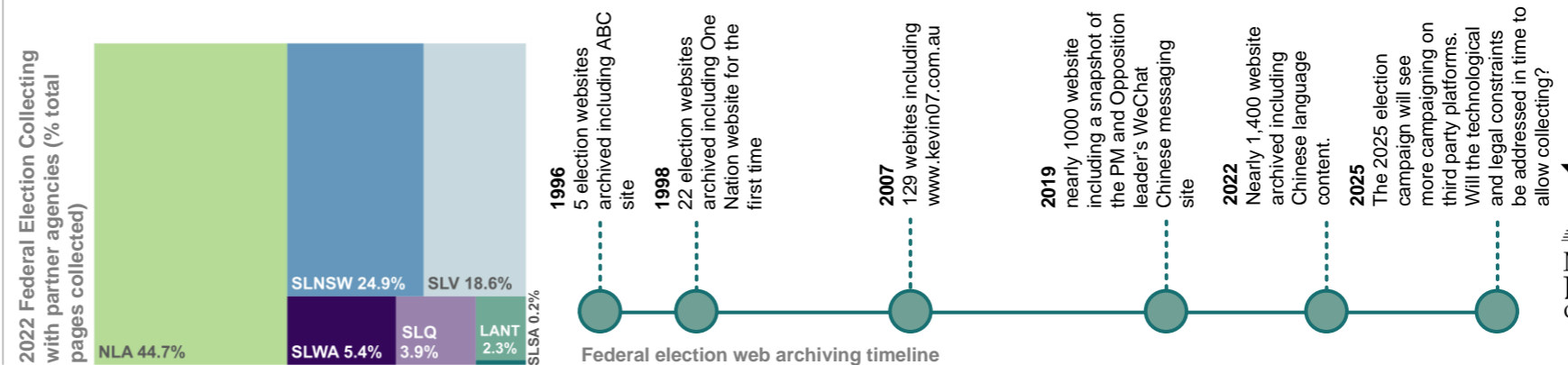
- The annual whole of domain harvest is an essential collecting activity, capturing websites that can disappear.
- The 2022 whole of .au domain crawl was completed, ingesting **82 TB** and **1 billion files**. An infrastructure upgrade at the Internet Archive meant that 2022 crawl was completed in half the time of previous years.
- In addition to the domain harvest, the Library conducted four bulk harvests of Commonwealth Government websites in 2021-22 adding more than 6 TBs of data or around 70 million files of government documents to the Archive.

Collaborative Web Archiving

- The Library has continued to provide gratis access to collaborative web archiving infrastructure since 1996. Tools developed by the Library have been used by State and Territory libraries, the War Memorial, the National Gallery and AIATSIS to collaboratively curate Australia's online heritage.
- The Library has continued its collaborative curated and thematic web archiving program working with State and Territory libraries to collect approximately 18 TBs of data or 140 million files from over 42,000 curated harvested instances in 2021-22.
- The web archiving partnership is an example of collection building that is collaborative and cooperative with shared goals and resources. The nation benefits from joint collecting on diverse topics such as COVID-19, bushfires and Federal elections.

2022 FEDERAL ELECTION

- Collecting political ephemera is a part of the National Library's role to collect, preserve and share the history of Australia. The collection has evolved over time from purely print related material to include digital material.
- In 2022 nearly 1,400 federal election related websites were collected in collaboration with web archiving partner agencies. Some sites were archived on multiple occasions to capture content changes throughout the campaign.



HOW WELL ARE WE DOING?

We have work to do to capture more comprehensively the online lives of Australians



FUTURE OPPORTUNITIES

Content

The web is the most dynamic medium of our time and if we are to collect online content that adequately reflects Australian online publishing and discourse, we require infrastructure that uses the best available tools.

Infrastructure

- The Library is committed to collaborative web archiving through:
 - Co-investment from partners in tools and infrastructure.
 - Co-investment will allow the Library and partners to continue to capture and preserve the web that exists now and what may be developed in the future.
 - Co-investment will deliver new generation collecting tools to enhance collecting of dynamic and social media content.

Research(er) Opportunities

- The Australian Web Archive as a historical data source provides an untapped resource for Library scholars and fellows.
- The Australian Research Data Commons and its work to build national digital research infrastructure provides opportunities for the Library to offer new tools and platforms to extend researcher capacity.

- The collection is discoverable through **Trove** and includes websites for political parties and candidates, polling, research, news and commentary as well as interest and lobby group websites.
- Selected snapshots of Twitter accounts were archived, including the Twitter accounts of each of the Teals.
- Large scale collecting of social media sites such as Facebook, Instagram, youtube and TikTok did not occur due to both legal and technical constraints.

