

The Australian Government Web Archive

ALIA Conference 2014

18 September 2014, Melbourne

Alison Dellit

Director, Australian Collection Management



NLA web archive collections

- PANDORA Archive collection (open access)
 - Selective web archiving since 1996
- Australian domain harvest collection (closed)
 - Large scale, outsourced (IA), annual collection, since 2005
- **Australian Government Web Archive** collection (open access)
 - Bulk seed list harvesting, outsourced (IA) and in-house run, annual (or more frequent)
 - 2011, 2012, 2013 (x2) and 2014 (x2)





So where did AGWA come from?

- Administrative conditions
 - Whole-of-Government arrangements
 - Gershon Review (Oct. 2008)
 - May 2010 –Secretaries’ ICT Governance Board approval
 - Non-corporate PGPA Agencies ✓
Commonwealth corporate entities ✗
- Technical and development considerations
 - NLA development of infrastructure and skills
 - Large scale, bulk harvesting
 - Access to large scale, bulk harvested collections

Selective

'targets', 'titles'

Themed

Curated seed lists

e.g. gov.au

2nd L Domain

e.g. *org.au*

TL Domain

i.e. *.au*

Whole Web

Internet Archive

Small scale

Moderate scale

Moderate to large scale

Large scale

Large scale

Reactive
Timely
Scheduled

Scheduled
Timely

Scheduled
(moderate control)

Scheduled
(low control)

Ongoing
Unscheduled

High curation

High curation

Moderate curation

Low curation

No curation control

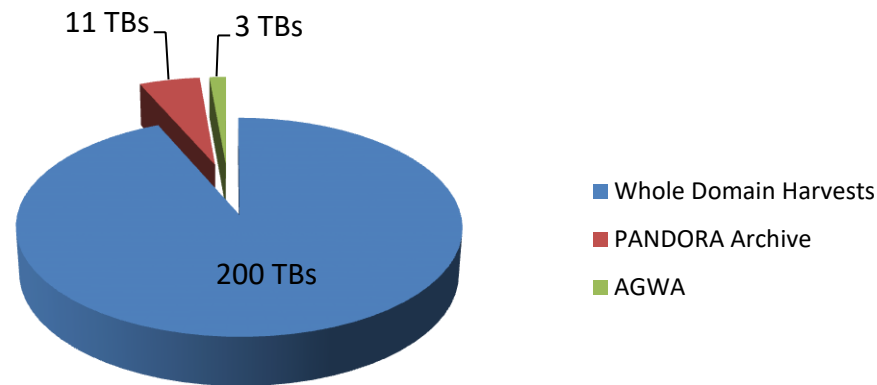
NLA Web Archiving Statistics

	PANDORA Web Archive	Australian Domain (.au) Web Archive	Australian Government Web Archive	All Collections
	‘Selective’	‘Country TL domain’	‘Seed-list’	
	1996 – Sept. 2014	2005-2014	2011-2014	
	(102,000 instances)	(9 crawls)	(6 crawls)	
Files	269 million	6.33 billion	76.9 million	6.67 billion
Data	13 TB	236 TB	7 TB	256 TB

AGWA content

	Total	Average harvest
Files	34.5 million	~ 8 million
Data	3 TB	750 GB – 1 TB

Data (TBs)



<http://webarchive.nla.gov.au/gov>



AGWA futures

Coming soon:

- 2005-2011 harvest content
- More commonwealth agencies
- More integration to a catalogue near you.

Next few years:

- Integration into Trove
- Metadata extraction
- Visualisation of data

<http://webarchive.nla.gov.au/gov>

Feedback to:

agwa@nla.gov.au

webarchive@nla.gov.au

