National
and State
Libraries
Australasia

NSLA Contemporary Indigenous Collections Audit

Stage 1: Published items, web archives and ephemera (full report)
JULY 2021

Table of Contents

Introduction

This project was designed to address a lack of data about the nature, description and discoverability of contemporary Indigenous collections in NSLA libraries.

NSLA libraries know that we have many improvements to make in the way that Indigenous collections are selected, described, managed, and made accessible. This has been acknowledged right across the cultural sector, as the movement toward cultural capability in cultural institutions has gathered pace. NSLA members have been collaborating on Indigenous collecting policies and projects for over a decade, including considerable investment in the Culturally Safe Libraries Program from 2018. What we have been lacking in this work is comparative data that can quantify and specify the problem.

In September 2020, four NSLA libraries embarked upon a research project to address this. By analysing a sample of published collections, web archives and original materials, the project would seek to answer questions such as: How do we make collecting decisions for contemporary Indigenous collections? Whose voices are really represented in those collections? What is the proportion of our collections created *by* Indigenous people versus *about* Indigenous people? How are our description practices affecting discovery of these collections? What are the major gaps in terms of geography, community representation, and format? How has our collecting practice for contemporary Indigenous materials evolved?

The first stage of the project would be a quantitative audit based on a sample of published items, web archives and ephemera. The second stage would be a qualitative analysis of a sample of original collections (manuscripts, photographs, oral histories, music or maps), considering the nature of community engagement in their acquisition, management, presentation and discoverability. Participating libraries are the National Library of Australia, State Library of Queensland, State Library of Western Australia, and State Library of New South Wales.

Scope

This report presents findings from **stage one** of the project. It includes:

- **Published collections, 2015-2020.** Analysis of published items identified as containing Indigenous material and catalogued – or with catalogue records enhanced – between 2015 and 2020.
- **Web collections, 1990s-2020**. Analysis of websites collected through the Australian Web Archive and PANDAS infrastructure, and social media collected by the State Library of New South Wales, about four specific national Indigenous campaigns.
- **Ephemera, 1990s-2020.** Analysis of ephemera collected by participating libraries relating to the same national campaigns as for web collections. Ephemera was selected for inclusion in this analysis as one of the formats most likely to be collected spontaneously in response to activist or community-led events.

This report presents only the 'high level' combined results of the data analysis. A large part of the project's value has been in the generation of very detailed datasets that can be made available to NSLA libraries for deeper interrogation, and the results are best considered alongside methodology notes from the data analysts (see report appendices for details).

Findings: summary

Published items

- A total of **18,398 published items** were identified with Indigenous content or references across the four participating libraries in this five-year period.

- Just under 60% of this published collection is in physical format, with 40% digital. However, 55% of collection records include a link to an electronic resource and 8% are included in NED.

- A ten-year trend analysis shows that digital collecting for this subject area peaked in 2018 and has since declined.

- The number one **record type** is language material, followed by non-projectable graphic and non-musical sound recording. The most common **bibliographic format** is monograph, and the most common **carrier type** is online resource.

- Of the total published items, 14,397 (or 78%) were **unique holdings**, leaving 22% or roughly one fifth of the collection as duplicate holdings.

- Between 14-19% of the published collection, or about one sixth, could be shown to have **Indigenous authorship** based on available external databases (combination of Indigenous authors on AustLit register and Aboriginal Land Council names).

- Of the top thirty **publishers** of Indigenous content, six are Indigenous owned or operate primarily for the publication of Indigenous material.

- Approximately one third – or 32% of records – for published items with Indigenous content include at least one **Indigenous language** listed in the AIATSIS Austlang thesaurus; and 30% include at least one **Indigenous place name** from the corresponding AIATSIS thesaurus.

- Of records including **metadata tags for AIATSIS thesauri**, only 6% included "aiatsisl" (language); 12% included "aiatsisp" (place name); and 12% included "aiatsiss" (subject).

- Records for 14% of the published collection include **access conditions**. Just 1% include **cultural sensitivity messages**.

Web archives, social media and ephemera

- De-duplicated samples taken from one year per decade (90, 2000s, 2010s) to show **growth in collections** for each of the four nominated national campaigns (NAIDOC Week, National Sorry Day, Mabo Day, Black Lives Matter) revealed:

    o National Sorry Day jumped from 91 sites in the 90s sample to 5,377 in the 2010s
    o NAIDOC Week jumped from 46 sites in the 90s sample to 3,580 in the 2010s
    o Mabo Day had a less dramatic rise from 52 in the 90s sample to 217 in the 2010s
    o Black Lives Matter does not show up until 2014 (28 sites), jumping to 4,445 in 2020

- The combined data for the Indigenous events and activism identified in the web archive shows 21%, or one fifth, being from **government domains**.

- Earlier periods of web collecting revealed very strong representation of **government and educational** websites/voices, with a greater proportion of dot com and dot org websites in later years suggesting more general **community online engagement**. However, it is known that many Indigenous organisations are not using the .au domain at all.

- **Public activism** as represented by Black Lives Matter has negligible content from government domains (0.2% in the 2020 dataset).

- The lack of structured or controlled metadata relating to all the resources in the data sets make it very difficult to determine the **voices** represented.

- **Curated collections** in Pandora are an effective way to collate web materials on a theme, and are accessible to browse through Trove, but are not well utilised or readily discoverable.

- Of active **social media queries** collected by SLNSW, 8% relate to Indigenous matters.

- Of the four nominated national campaigns, **Black Lives Matter** was by far the most strongly represented in the dataset. The majority of content comes from **Twitter**. Of the top ten **authors** and top ten mentioned authors in this dataset, just one was Indigenous-identified, and two were publicly supported by recognised Indigenous leaders.

- Physical **ephemera collections** relating to the four nominated national campaigns were patchy – in some cases non-existent – and difficult to identify using catalogue records. The majority of content came from government or official sources.

Metadata limitations

The limitation common to all collections considered in this audit was description and discoverability. Limitations of the metadata meant that a number of research questions could not be meaningfully answered. These included:

- What are the most common subject terms in collection records (e.g. art, sport, legal, biography)?

- What proportion of collection items were published outside of capital cities?

- How many items were actually published between 2015-2020 as opposed to having had their records enhanced in that period?

- Which/how many records for photographic items include generic descriptors like man, woman, child, family, native and black?

Results are presented in more detail in the next section of the report.

Part I: Published collections, 2015-2020

Methodology

The raw material for this piece of analysis was extracted by the Trove Collaborative Services team (NLA) from the Australian National Bibliographic Database (all records appear in Trove). The extraction was based on this query:

> *(AW:Indigenous AND AW:Australia*) OR (AW:Aborig* AND AW:Australia*) OR (AW:"Torres Strait Island*" AND AW:Australia*) OR AW:"First Australians" OR AW:"Aborigin* and Torres Strait Island* peoples" OR "Australian Languages") AND DEF:[20150101 TO 20201231]*

Additional records were identified by the presence of AIATSIS subject headings in the 600 MARC fields, or AUSTLANG codes in the 041 field. **The search yielded 18,398 results across the four participating libraries**.

A set of research questions was identified by the project group to be applied to this data, and a methodology proposed based on the content of the relevant MARC fields. Given the size and complexity of the dataset, a professional data analyst was contracted to undertake this work with assistance from the group.

One early and important finding was that a number of research questions *could not be pursued* due to metadata inconsistencies. See *Appendix C* for details of methodology and the MARC fields used in this analysis.

Results

UNIQUE HOLDINGS
Libraries wished to know the number of holdings in the dataset that were unique to them.
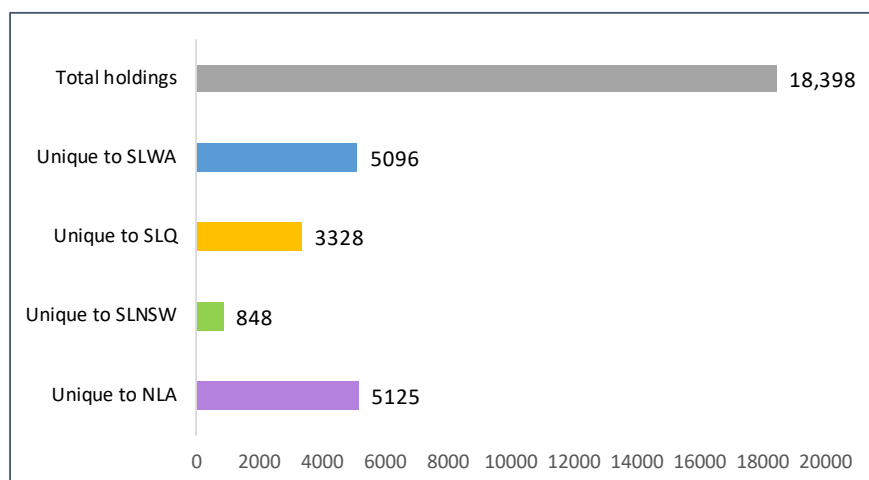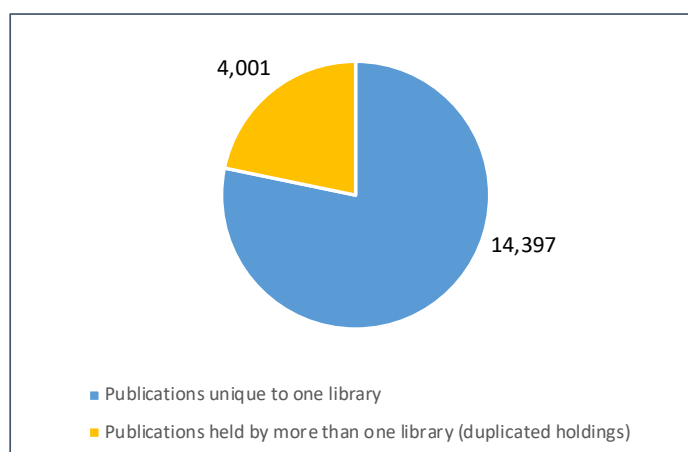

FIGURE 1: NUMBER OF UNIQUE HOLDINGS IN EACH LIBRARY (NOT HELD BY THE OTHER THREE PARTICIPATING LIBRARIES)

FIGURE 2: NUMBER OF PUBLICATIONS UNIQUE TO ONE LIBRARY (78%): AGAINST DUPLICATE HOLDINGS (22%)

AUTHORSHIP AND VOICE

Establishing Indigenous authorship would be impossible without investigating each individual title. Three options were instead put forward that could give us an *indication* of Indigenous authorship and voice. Two required cross-checking with external databases, whose quality and veracity were out of the project team's control.

The first exercise was to extract records that included the name of an Aboriginal Land Council, comparing with state-based lists.[1] The second was to extract records that included the name of authors registered with AustLit[2] as being of Aboriginal or Torres Strait Islander descent. The third was to extract a list of the most frequently represented publishers and to look at the proportion with Indigenous ownership and/or a focus on Indigenous audiences.
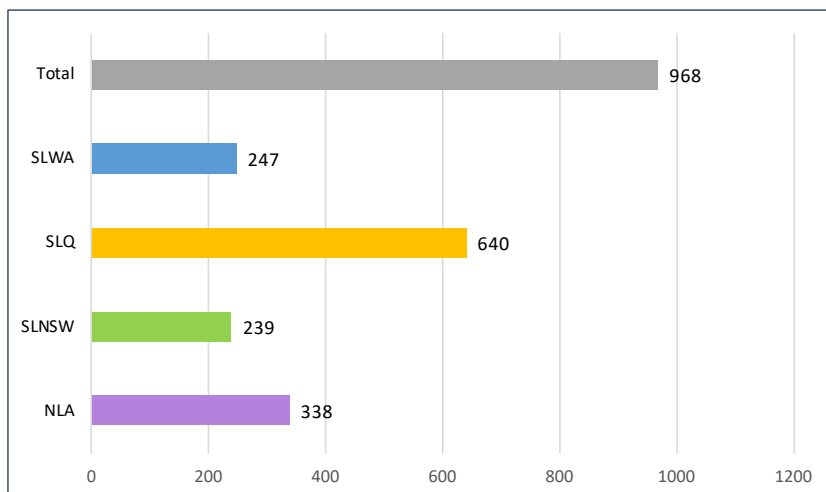

FIGURE 3: NUMBER OF PUBLICATIONS INCLUDING THE NAME OF AN ABORIGINAL LAND COUNCIL[3]: 968 OF TOTAL 18,398
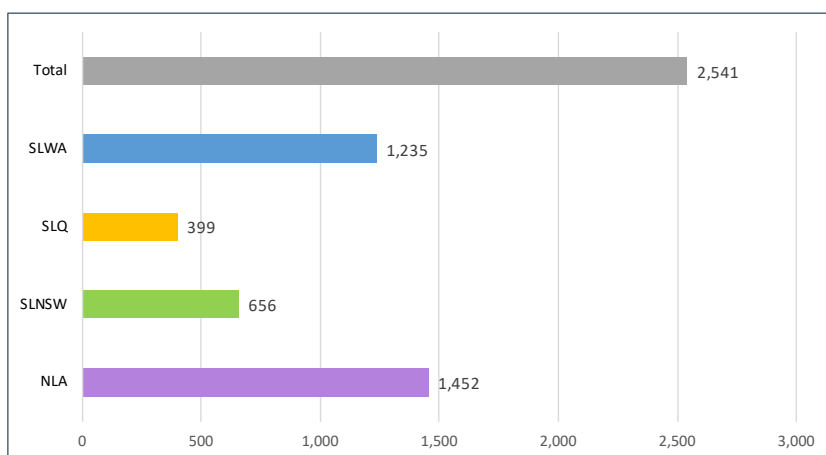

FIGURE 4: NUMBER OF PUBLICATIONS INCLUDING THE NAME OF AN ABORIGINAL OR TORRES STRAIT ISLANDER AUTHOR REGISTERED WITH AUSTLIT[4]: 2,541 OF TOTAL 18,398

---

[1] https://www.creativespirits.info/aboriginalculture/selfdetermination/aboriginal-land-councils#list-of-aboriginal-land-councils-in-australia

[2] https://www.austlit.edu.au/blackwords

[3] **Note on methodology**: Aboriginal Land Councils rarely appear with their full names described in the metadata. For instance, "Walgett Local Aboriginal Land Council" doesn't appear in the dataset, but some titles include "Walgett Aboriginal Medical Service Co-operative". I thus simplified the list of Aboriginal Land Councils. For instance, we looked for the appearance of the words 'Walgett', or 'Winbar' in the given MARC fields ('100-MARC','110-MARC','111-MARC','600-MARC','610-MARC','260-MARC','264-MARC','542-MARC','700-MARC'). As it wasn't clear which MARC field was for Aboriginal Land Council or for Aboriginal authors, I used the same for both. Datasets provided to project members include a final column containing the Aboriginal Land Council name that was matched against the record.

[4] **Notes on methodology**: Numerous names in the AusLit list where listing only a first name and many were very common Anglo first names. To avoid too many false positives, I filtered the AusLit list down to only include non-anglo
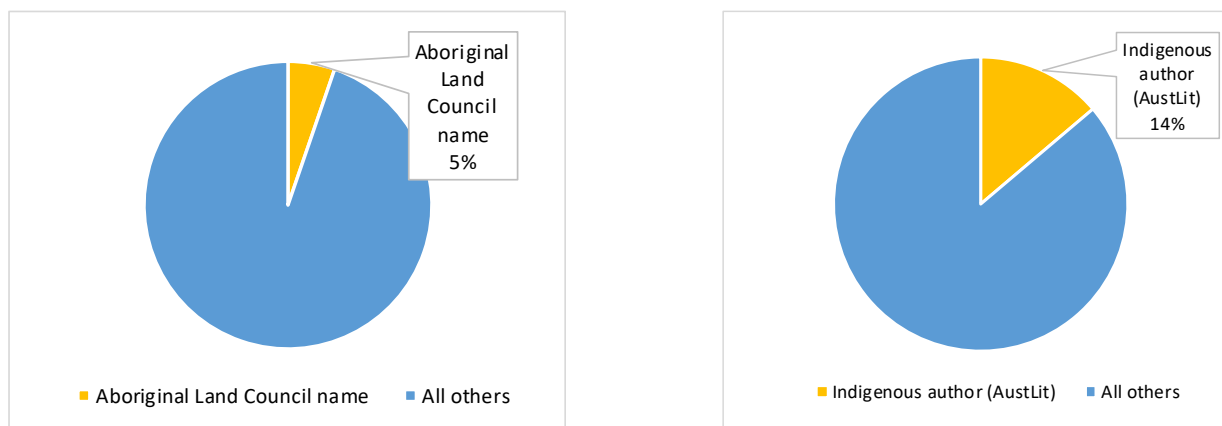
FIGURES 4 & 5: PERCENTAGE OF RECORDS INCLUDING THE NAME OF AN ABORIGINAL LAND COUNCIL; AND PERCENTAGE OF RECORDS INCLUDING THE NAME OF AN AUSTLIT-REGISTERED AUTHOR: AGAINST TOTAL RECORDS

Notwithstanding the difficulties with this measurement (see footnotes) – and recognising that the AustLit register will be far from comprehensive – the indication is that Indigenous-identified authorship is low in a contemporary collection of published materials with Indigenous subject matter. This is especially marked considering the number of unique holdings across the four libraries.

In addition, SLQ has noted the important caveat that 440 of the records marked as including an Aboriginal Land Council for that state, and 399 of those including an Aboriginal author, are in fact digital photographs, which the library has considered has published rather than 'original' materials since 2018. Those photographs on the list can be attributed to five individual photographers.

A table of the top thirty publishers appearing in the collection – based on data extracted from the MARC 264 field – appears below. Those with Indigenous ownership or Indigenous subject matter expertise are marked in grey.

	Publisher	Records		Publisher	Records
1	Australian Institute of Health and Welfare	212	16	Cambridge University Press	60
2	Routledge	176	17	ACT Government	58
3	Allen & Unwin	139	18	ETT Imprint	57
4	Magabala Books	116	19	Redback Publishing	47
5	Batchelor Press	94	20	Australian Indigenous HealthInfoNet	44
6	ReadHowYouWant.com	84	21	Indigenous Literacy Foundation	44
7	Palgrave MacMillan	83	22	NewSouth Publishing	43
8	Bolinda Audio	77	23	HarperCollins Publishers	42
9	Hesperian Press	72	24	CSIRO Publishing	41
10	Aboriginal Studies Press	65	25	Penguin Random House Australia	41
11	ANU Press	64	26	Kimberley Aboriginal Medical Service	37
12	Knowledge Books and Software	64	27	Wakefield Press	35
13	UWA Publishing	63	28	University of Queensland Press	35
14	Oxford University Press	61	29	Fremantle Press	34
15	Hachette Australia	61	30	Springer	33

TABLE 1: LIST OF TOP THIRTY PUBLISHERS REPRESENTED IN THE COLLECTION, WITH INDIGENOUS OWNERSHIP OR SUBJECT MATTER EXPERTISE HIGHLIGHTED IN GREY.

first names when only a first name was provided. To do this, I removed the top-1000 names registered in Scotland from 1880-2008. All other names were kept.

DESCRIPTION

A particular focus of this piece of analysis was the degree to which libraries are using AIATSIS thesauri for Indigenous languages, place names and subject terms. These records are ideally tagged in publication metadata with "aiatsisl", "aiatsisp", and "aiatsiss" respectively in the MARC 650 field.

Results showed that 5,904 records included at least one Indigenous language from the AIATSIS Austlang thesaurus.[5] This represents 32%, or about one third, of the total number of records.
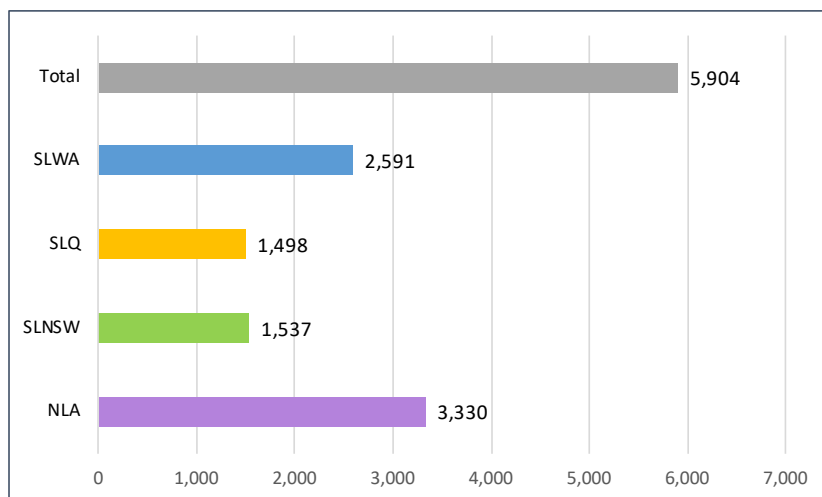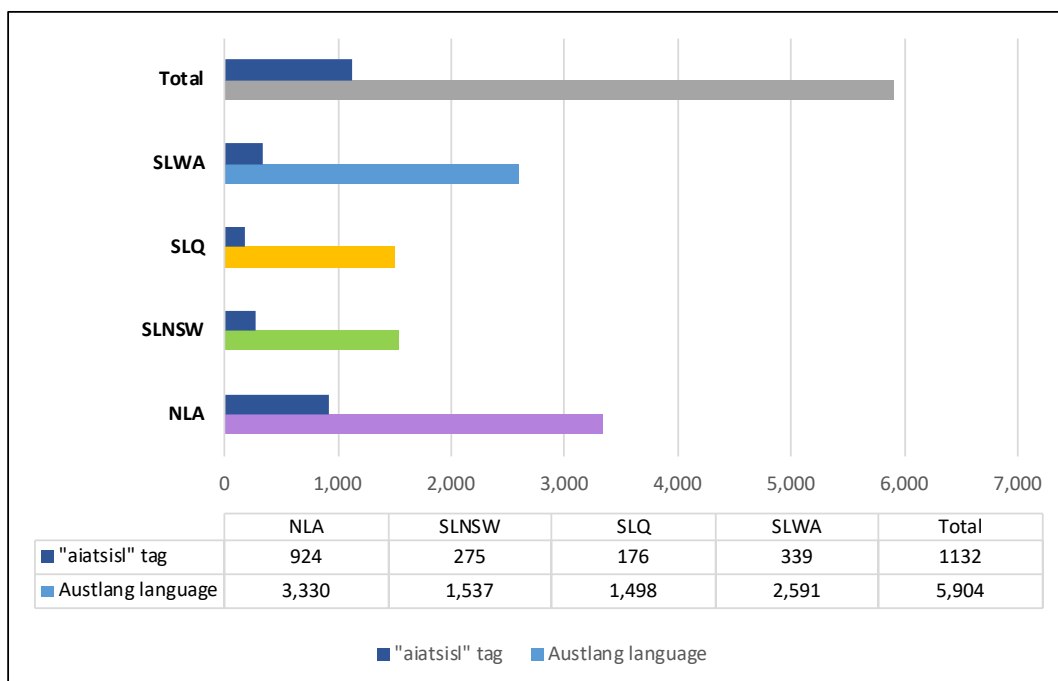

FIGURE 6: NUMBER OF RECORDS THAT INCLUDE AT LEAST ONE **INDIGENOUS LANGUAGE** FROM THE AUSTLANG THESAURUS (COMPRISING 32% OF TOTAL NUMBER OF RECORDS)

The number of records including an Indigenous language listed in the Austlang thesaurus, however, did not correspond with the number of records in which the "aiatsisl" tag appeared: the tag designed to signify use of the thesaurus in a record. This number was considerably lower.

FIGURE 7: NUMBER OF RECORDS INCLUDING AN AUSTLANG INDIGENOUS LANGUAGE: AGAINST THOSE USING THE "AIATSISL" TAG

	NLA	SLNSW	SLQ	SLWA	Total
"aiatsisl" tag	924	275	176	339	1132
Austlang language	3,330	1,537	1,498	2,591	5,904

■ "aiatsisl" tag   ■ Austlang language

---

[5] **Note on methodology**. I used both 'language name' and 'language synonym' columns in the AustLang dataset. Both had often lists that I prefiltered to remove common terms that might not designate a language so as to avoid false positives (eg 'southern', 'north', 'mount'). 11,504 terms were kept to designate languages. The list is provided for reference. I used the following MARC fields: '008-MARC','041-MARC','546-MARC', as well as all '650-MARC' fields.

A similar result was found when the records were searched for Indigenous place names from the AIATSIS thesaurus as against use of the "aiatsisp" tag.[6] Place names had been assigned to 30% of the collection.

	NLA	SLNSW	SLQ	SLWA	Total
■ "aiatsisp" tag	1,514	598	843	488	2,266
■ AIATSIS place name	2,104	737	2,639	1,484	5,503

■ "aiatsisp" tag   ■ AIATSIS place name

FIGURE 8: NUMBER OF RECORDS INCLUDING AN INDIGENOUS PLACE NAME IN THE AIATSIS THESAURUS: AGAINST THOSE USING THE "AIATSISP" TAG
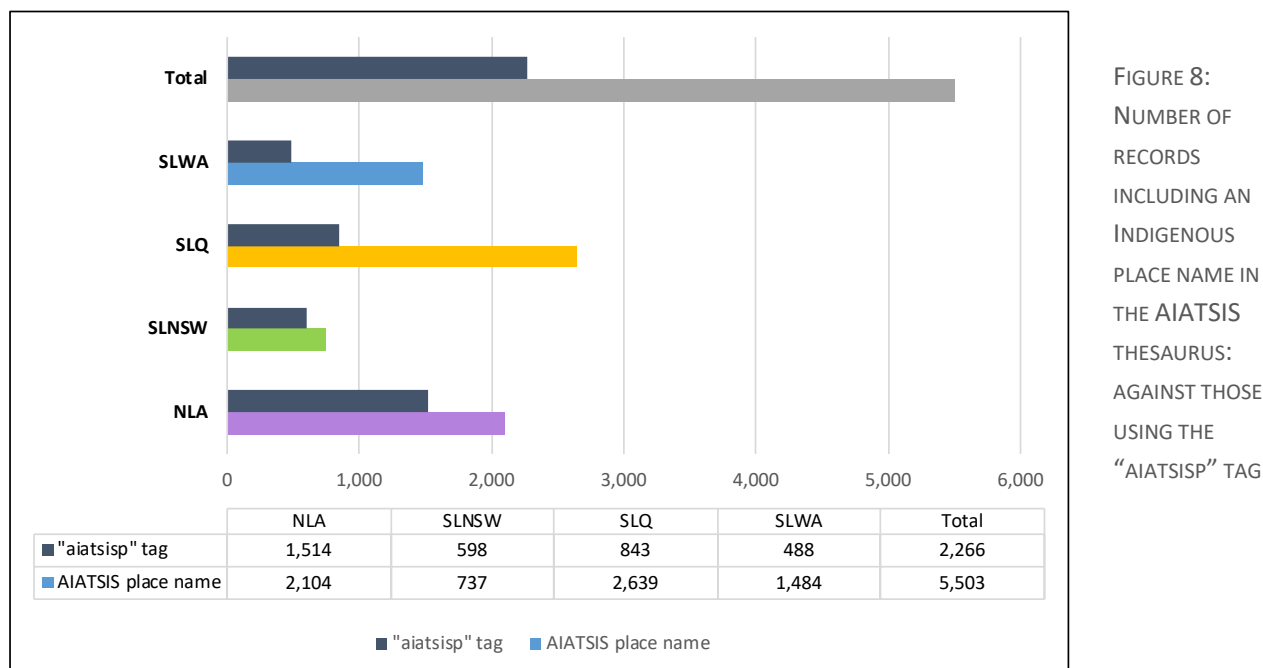
A search of individual Indigenous subject terms in the AIATSIS subject thesaurus was not completed, as the list includes a number of very broad terms which are difficult to extract meaningfully (without the specificity of languages or place names). The number of "aiatsiss" tags used in the collection, however, was calculated at 2,255 – or 12% of the total collection.

These numbers can be explained in some cases by the relatively short length of time that libraries have been formally using the thesauri – many libraries have been using the Austlang codes for less than two years. SLWA has been including relevant language and subject coding in all records done in-house since 2019, but it was noted that this is not always done consistently and those records missing information are not always followed up. Retrospective record enhancement depends on staff time and resources, and in some cases, the recommendations provided by the thesauri have not been followed. SLQ noted that it uses the 'aiatsisl' tag to indicate First Nations language material more generally, with or without the specific languages included.

---

[6] **Note on methodology:** All terms were extracted from the place names thesaurus and any English place names or overseas cities and countries were removed before carrying out the analysis.

FORMAT

Project members were interested to look at collection formats, starting with physical versus digital, number of holdings in NED, digital collecting trends, and most common formats for Indigenous collections overall.
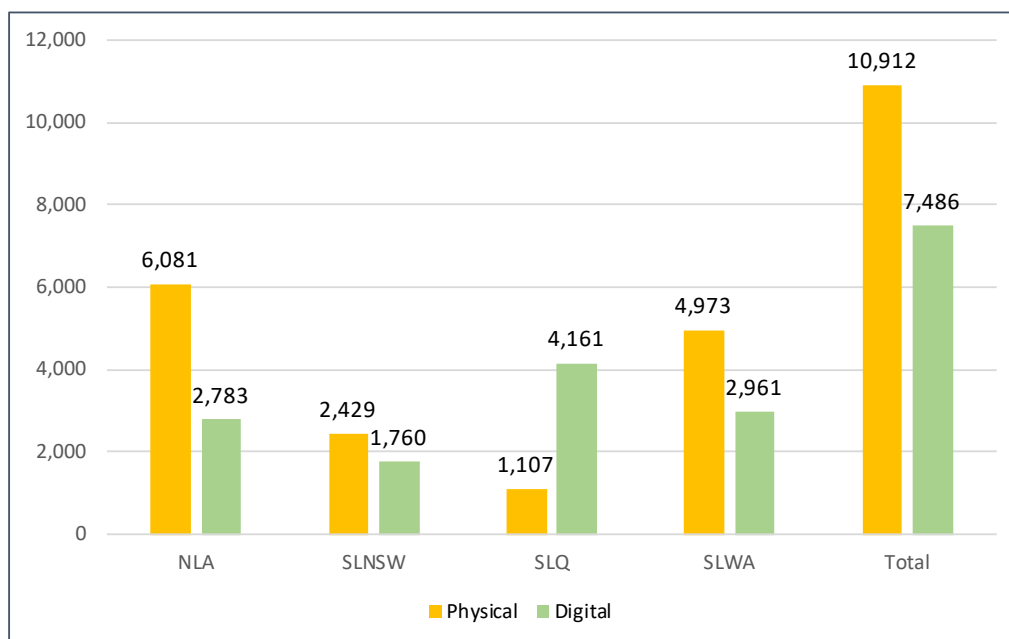
FIGURE 9: NUMBER OF PHYSICAL VERSUS NUMBER OF DIGITAL ITEMS IN THE COLLECTION

SLQ's collection comprises significantly more digital material than other states, which reflects the weighting of its acquisition policy toward digital collecting over physical collecting. SLWA noted that its physical holdings still outweigh digital ones, suggesting this may reflect collecting practices and the limitations of systems to manage digital formats and processes effectively. An analysis across all four libraries found that digital collecting peaked in 2018, as in *Figure 10* below.

FIGURE 10: TEN-YEAR DIGITAL COLLECTING TREND ACROSS ALL FOUR PARTICIPATING LIBRARIES
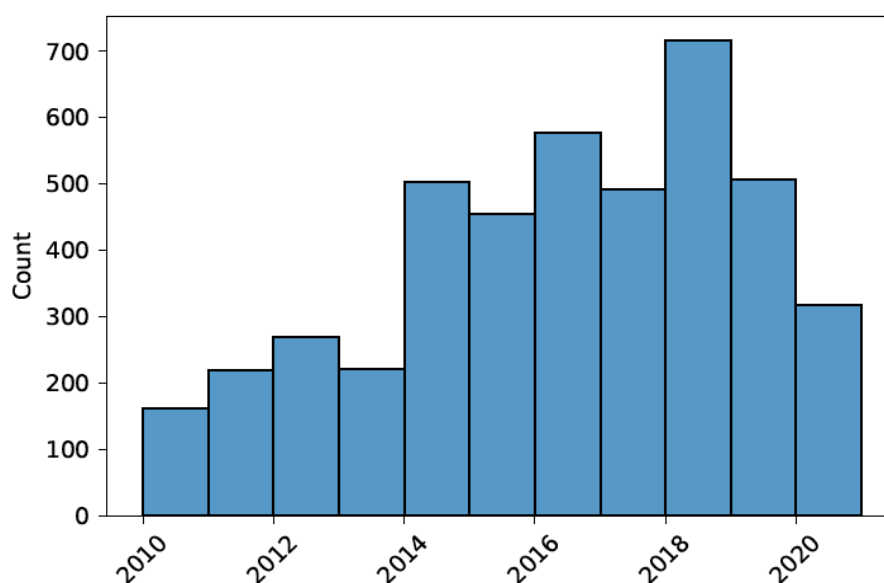
The proportion of records included in the National eDeposit Service (NED), and the number containing a URL (i.e. linked to an eresource) were 8% (1,430) and 55% (10,181) respectively.
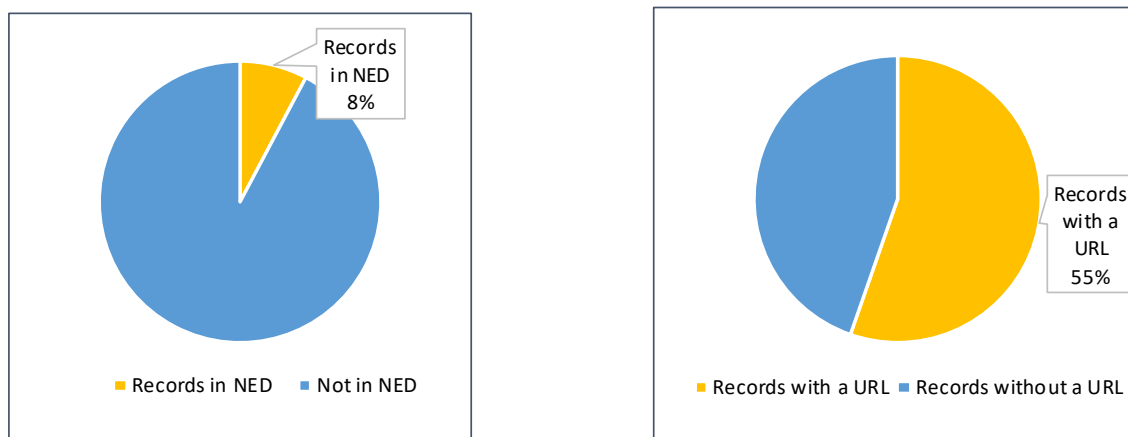
FIGURES 11 & 12: PERCENTAGE OF RECORDS HELD IN NED AND PERCENTAGE CONTAINING A URL VERSUS TOTAL NUMBER OF RECORDS.

The tables below list the number of records by record type, bibliographic level, and carrier type. Provided here are the top twenty carriers but a full list is available.

Record type	Number of records
Language material	8,789
Two-dimensional non-projectable graphic	7,429
Non-musical sound recording	758
Projected medium	731
Cartographic material	205
Manuscript language material	127
Musical sound recording	115
Computer file	88
Notated music	68
Manuscript notated music	27
Mixed materials	23
Three-dimensional artefact or naturally occurring object	22
Kit	15
Manuscript cartographic material	1

TABLE 2: DISTRIBUTION OF RECORDS ACCORDING TO RECORD TYPE (MARC 300: PHYSICAL DESCRIPTION)

Bibliographic format	Number of records
Monograph/item	12,568
Subunit	4,536
Collection	466
Integrating resource	397
Serial	292
Monographic component part	138
Serial component part	1

TABLE 3: DISTRIBUTION OF RECORDS ACCORDING TO BIBLIOGRAPHIC FORMAT (MARC 06 & 07)

Carrier	Number of records
online resource	4,829
volume nc	3,143
volume	3,004
sheet	2,162
online resource cr	1,187
slide	1,013
other nz	370
videodisc vd	300
audiocassette	273
slides	199
audio disc sd	185
sheet nb	136
other	118
videocassette	88
audiocassette ss	63
text	45
videodisc	44
volume nc	43
videocassette	43
audiotape reel	39

TABLE 4: TOP TWENTY CARRIER TYPES REPRESENTED IN THE COLLECTION [CATEGORIES EXACTLY AS THEY APPEAR IN THE RECORDS] (MARC 338)

ACCESS

The analysis of formats and digital materials above tells us something about access. The majority of NED holdings are openly available, and more than half the collection has a digital form or component. But how accessible is the collection overall? And does digital necessarily equate to access? A couple of additional extractions were run to look at the number of records containing access conditions and those containing cultural sensitivity messages.

Access conditions included 'not for inter-library loan', 'not for loan', 'not for general viewing', 'restrictions', 'restricted', 'available onsite' and 'access closed', included in the MARC 506 field.

Total records 18,398
Total access conditions 2,536
SLWA 1,661
SLQ 1,557
SLNSW 1,412
NLA 2,062
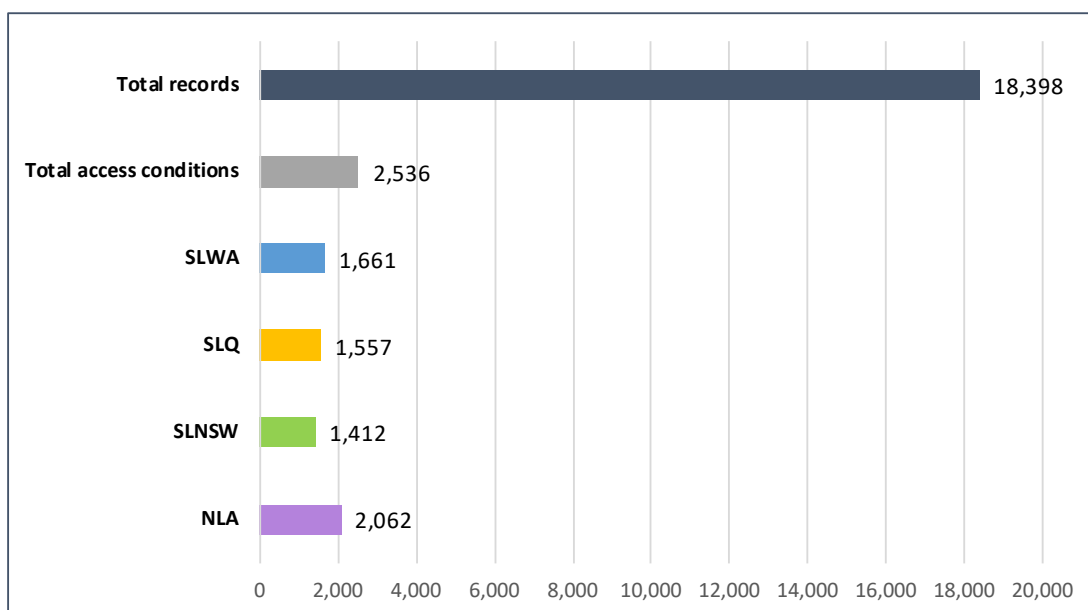
FIGURE 13: NUMBER OF RECORDS WITH ACCESS CONDITIONS: TOTAL AND BY INDIVIDUAL LIBRARY (14% OF TOTAL RECORDS)

FIGURE 14: NUMBER OF RECORDS CONTAINING CULTURAL SENSITIVITY MESSAGES (1% OF TOTAL RECORDS)
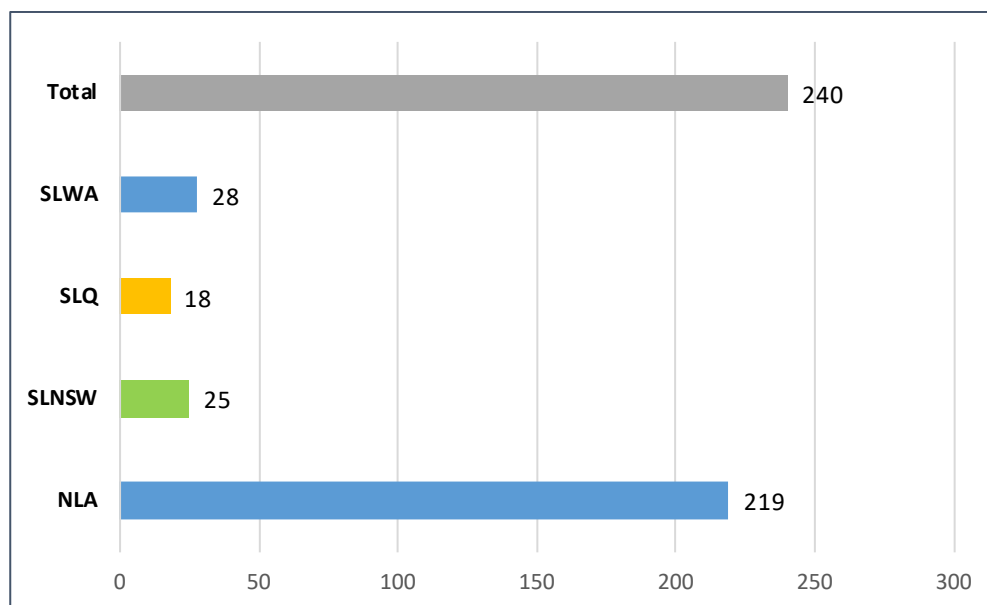
The number of records containing cultural sensitivity messages was obtained by extracting those with the words 'sacred', 'ceremonial' or 'deceased' in the MARC 500 field. This number (240) comprised just 1% of the total number of records.

*****

Part II: Web archives and ephemera, 2000-2020

Methodology

For this piece of analysis, four national Indigenous events or activist campaigns were selected by the group. Based on sampling over a twenty-year period, the group looked at the extent to which these events were reflected in the Australian Web Archive, the Pandora selective web archive, the State Library of New South Wales' social media collection, and ephemera collections across all libraries. Again the focus was on the range of content, its accessibility, and the voices represented.

For a detailed explanation of the methodology used in this analysis, please see *Appendix B*. The four events were: **National Sorry Day/Reconciliation Day, Mabo Day, NAIDOC Week, and Black Lives Matter**.

Analysis was also undertaken in relation to one local campaign identified by each of the participating libraries. This was to ensure that the exercise had some local relevance and application. Results for these local events are not included in this paper, but can be supplied on request. The local events were: Reconciliation Day (ACT), Laura Quinkan Dance Festival (QLD), Mowanjum Festival (WA), and Koori Rugby League Knockout (NSW).

Australian Web Archive

Fifteen datasets were extracted for materials relating to the chosen events from the Trove Australian Web Archive (AWA). The AWA is full text indexed and is publicly searchable through the Trove discovery service. However, it is not possible to download datasets of search results from the public discovery service. For this reason, the Solr Admin application was used to run advanced searches with results which could then be downloaded. Some deduplication was applied to the datasets. Searches were based on the following terms:

1. **1998**, **2008, 2018**: National Sorry Day ["Sorry Day", "Stolen Generations", "National Day of Healing", "Bringing Them Home", "National Apology"]
2. **1996**, **2006, 2016**: NAIDOC Week ["NAIDOC", "NADOC" (former name), "NAIDOC Week", "National NAIDOC Committee"]
3. **2013** and **2020**: Black Lives Matter ["Black Lives Matter", "BLM"]
4. **1996**, **2006, 2016**: Mabo Day ["Mabo Day", "Eddie Mabo", "Mer Island"]

The datasets were then individually analysed to break down the content by the principal 2nd level domains, i.e. org.au, edu.au, gov.au, net.au, com.au and asn.au. A count was also done to identify non.au content.

PANDAS (PANDORA web archive)

PANDAS is the workflow system and database used by PANDORA Archive partner agencies. PANDORA is the curated selective component of the Australian Web Archive which has been operational since the late 1990s.

The Indigenous-related content collected through PANDORA is discoverable primarily based on subject listings. These subjects do not use any specific thesaurus but are specific (albeit generic) terms used for PANDORA. Subject listings cannot be browsed through the Trove discovery service but are accessible through the PANDORA public website (http://pandora.nla.gov.au/subject/12). They include the top-level listing 'Indigenous Australians' and 11 sub-categories including: Government Indigenous Policy; Indigenous Art; Indigenous Business & Commerce; Indigenous Culture; Indigenous Education; Indigenous Employment; Indigenous Health; Indigenous History; Indigenous Land Rights; Indigenous Languages; and, Indigenous Native Title.

A dataset of **3,108** titles consisting of PANDAS metadata was derived using Metabase in February 2021 for all listings under the subject 'Indigenous Australians' and the 11 Indigenous sub-categories.

Social media

Using the same search terms listed above, a search was run across nine years of social media collections (Twitter, Facebook, and Instagram) held by the State Library of New South Wales on the Vizie platform (2012-2021). Posts were tallied for 2020, 2017, and 2012. In 2012, the State Library was piloting the Vizie platform.

Ephemera

A search of physical ephemera was undertaken in each library based on the same search terms above, and listed by number of items or boxes. Some illustrative examples are provided in this report.

Australian Web Archive and PANDORA

COLLECTION SIZE AND DISTRIBUTION

The graph below displays the number of websites (de-duplicated) collected in the Australian Web Archive about the four nominated national Indigenous campaigns based on one sample year per decade. *Note:* data in 2020 was collected only for Black Lives Matter.
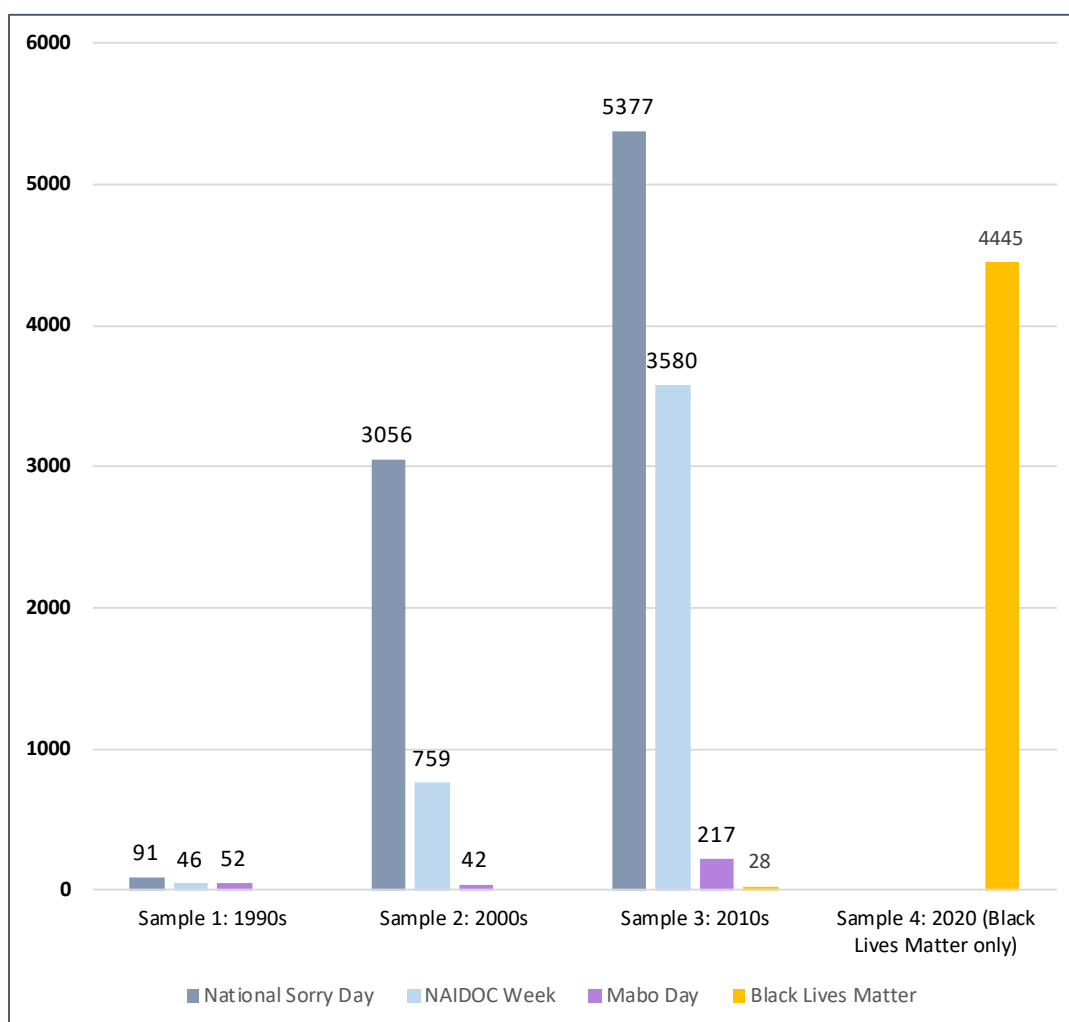

FIGURE 15: AUSTRALIAN WEB ARCHIVE COLLECTIONS ABOUT FOUR INDIGENOUS CAMPAIGNS BASED ON ONE SAMPLE YEAR PER DECADE

Collections in PANDORA on the same four campaigns over the full PANDAS dataset were modest, noting that these figures were based on a search of subject headings and URLs only:
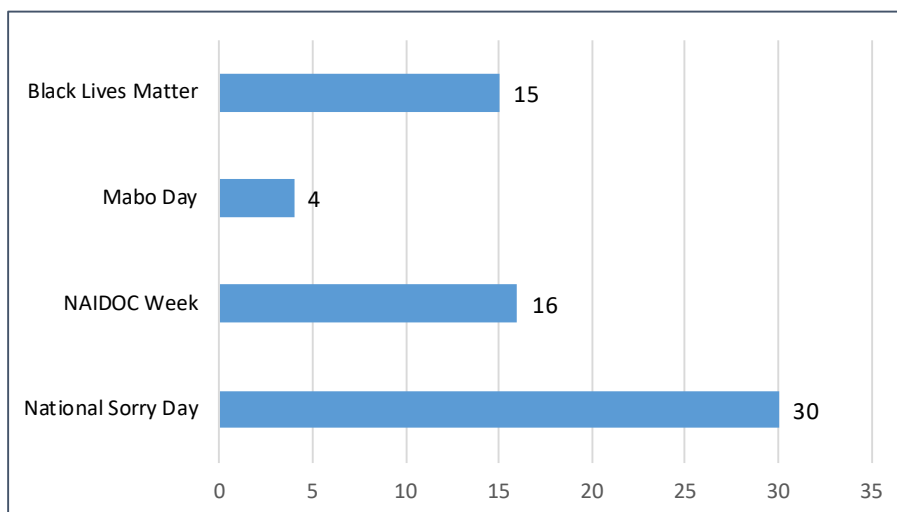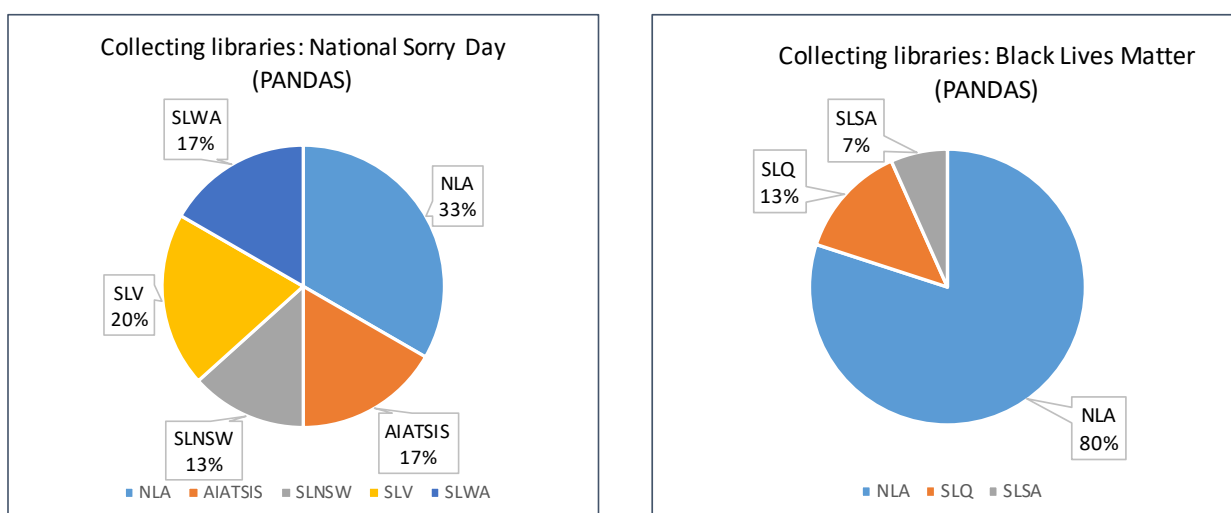

FIGURE 16: NUMBER OF SITES COLLECTED IN PANDORA (NOT TIME LIMITED) ABOUT FOUR INDIGENOUS CAMPAIGNS

The distribution of these collecting activities across PANDORA partners differs substantially between a longer-running campaign such as National Sorry Day and a recent one such as Black Lives Matter:

FIGURES 17 & 18: DISTRIBUTION OF COLLECTING ACTIVITY BETWEEN PANDORA MEMBERS FOR TWO CAMPAIGNS

The results tell us that these collections are difficult to reach based on a simple subject search, as search terms may not appear in the limited metadata fields available (i.e. seed URL, short title, subject). We know that there are more collections for Black Lives Matter, for example, as the subject of one of five curated collections of Indigenous materials included in PANDORA, and accessible to browse through Trove:

- Black Lives Matter Movement in Australia [80 titles; collaborative by NLA, SLSA, SLQ and SLWA]
- The National Apology [8 titles; collaborative by NLA, SLWA and AIATSIS]
- NITV News (SBS) [41 titles; an NLA collection; regular archiving of the NITV site from 2018 onwards]
- Victoria Treaty Advancement Commission videos [18 titles; an SLV collection of videos]
- Site identification mini posters [14 titles; an SLV collection]

AUTHORSHIP

In order to gain some sense of authorship – Indigenous or otherwise – for the Australian Web Archive sample, an analysis of second level domain type was undertaken.
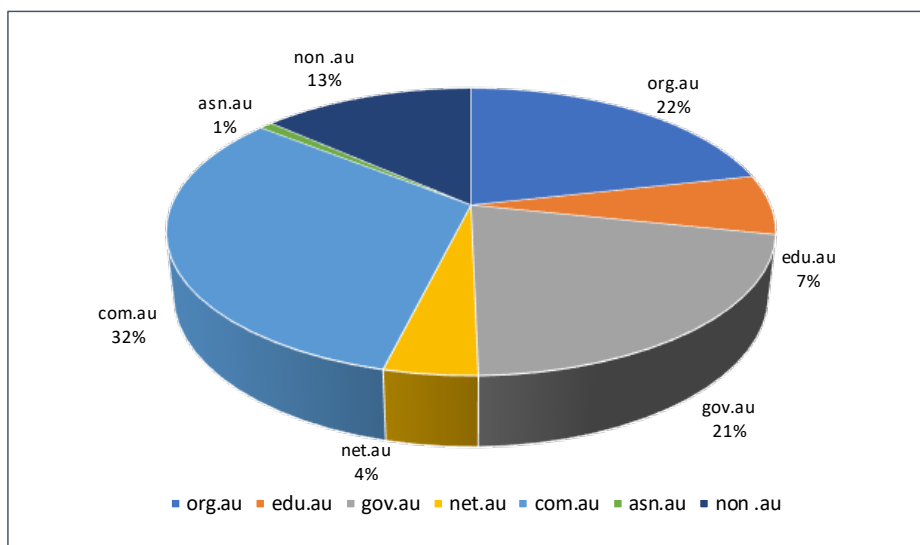
FIGURE 19: BREAKDOWN OF INDIGENOUS COLLECTIONS IN AWA BY 2ND LEVEL DOMAIN

The second level domain breakdown reveals strong representation of government and educational websites in the earlier period – representing official and research sources – while the .com.au (and to some extent the non .au) domain appear with greater prominence in later periods suggesting, perhaps, more general community online engagement. It is known, however, that many Indigenous organisations do not use the .au domain at all, with a preference for .net (e.g. indigenousarchives.net) and some social media platforms that cannot be collected through the AWA. Time-stamped snapshots for National Sorry Day and Black Lives Matter are presented below to demonstrate the evolution in content sources for web archives over time.

BLACK LIVES MATTER

A search was undertaken for sites relating to Black Lives Matter in 2013 (the year in which the movement was established) and 2020 (the year in which it received widespread international recognition). The charts below show a high proportion of non .au content collected in the small 2013 sample (28 records).
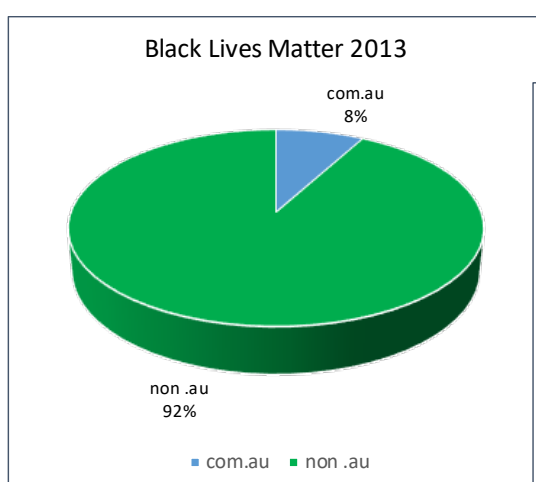
FIGURE 20: CONTENT SOURCES FOR AWA SITES RELATING TO BLACK LIVES MATTER IN 2013 AND 2020.
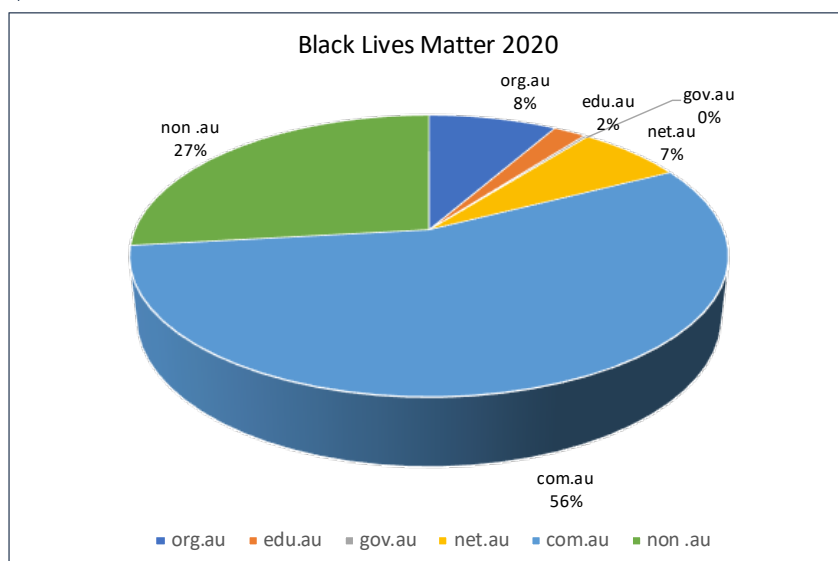
It should be noted that non .au content does not necessarily represent non-Australian content. This category includes many Australian targeted sites including news sites (such as guardian.com/au and buzzfeed.com), some scraped Twitter content and outlier domains such as .club. There will, however, be some non-Australian content included as a result of the scoping of domain harvests (prior to 2016) that included capturing embedded linked content, including non .au content (to the distance of 1 linked hop) for the purpose of context. This is the case for the 2013 [2014] Black Lives Matter dataset, which is primarily scraped tumblr.com and facebook.com pages and reflects the fact that, at that time, Black Lives Matter discussion was not appearing on the Australian web domain.

NATIONAL SORRY DAY

The charts below show an evolution over twenty years from a prevalence of educational sites in the 1990s to government sites in the 2000s, and subsequent increase in the output of organisations using the org.au domain.

FIGURE 21: EVOLUTION IN 2ND LEVEL DOMAINS REPRESENTED IN THE AUSTRALIAN WEB ARCHIVE OVER TIME FOR COLLECTIONS RELATING TO NATIONAL SORRY DAY

Broadly speaking, the web archive data confirms that online discourse around Indigenous activism and events has existed from the earliest times of online publication. The combined data for the Indigenous events and activism identified in the web archive shows 21% being from government domains.

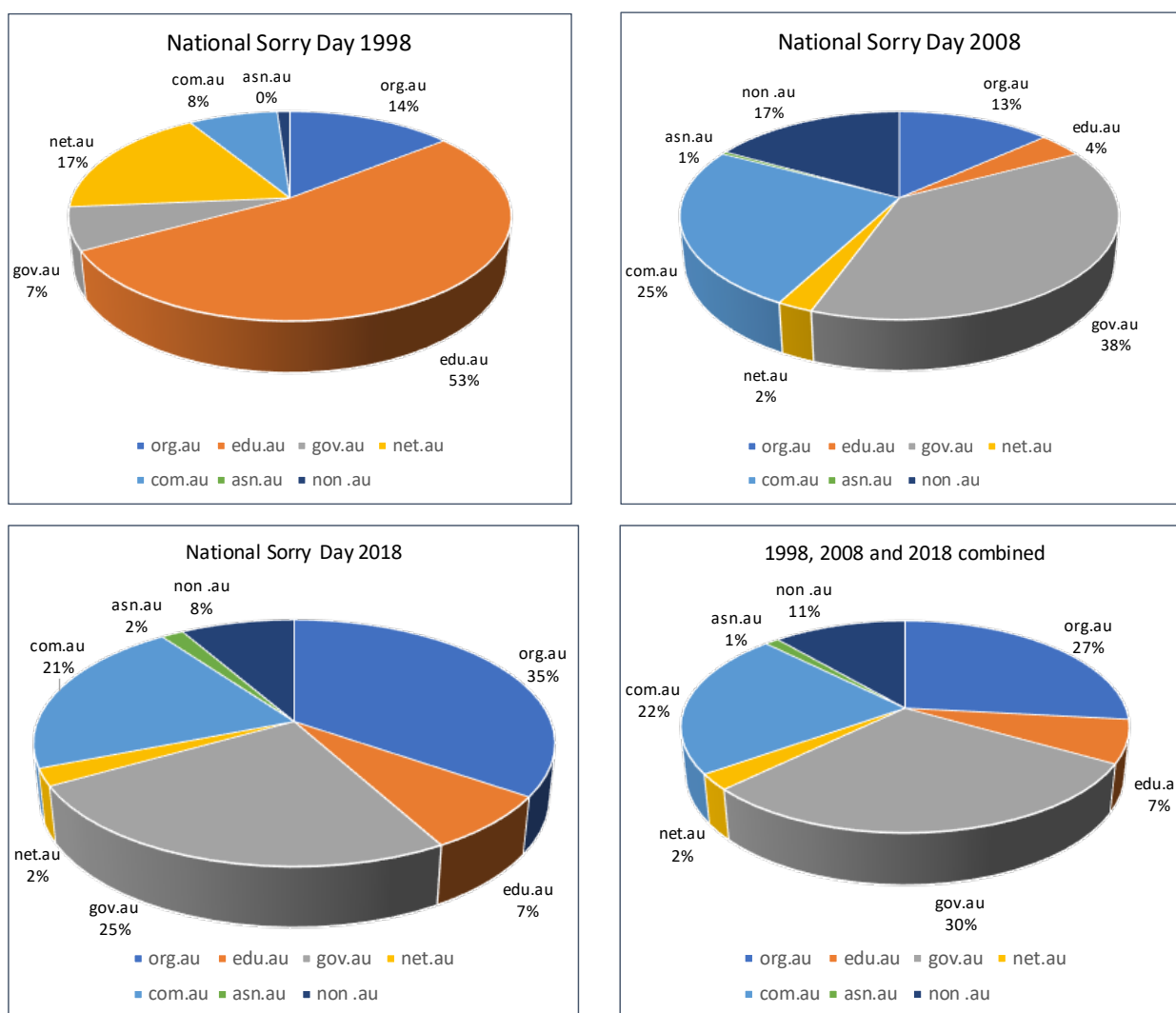Sanctioned events such as NAIDOC demonstrate a higher source of content from government domains in earlier years (59% in 1996 and 57% in 2006, less in later years down to 19% in 2016). Public activism as represented by Black Lives Matter has negligible content from government domains (0.2% in the 2020 dataset).

Looked at over the entire period, the .com.au, .gov.au and .org.au domains occupy the largest share of content. The .edu.au domain is somewhat less represented in the data (6% overall), however web archives content consists of freely available online Australian published content only and is unlikely to include material published in scholarly journals or content held in institutional repositories. There is some indication that .com.au and non .au domains increased in prominence in more recent times, perhaps suggesting increasing community engagement online supported by more readily available publication platforms. While a quarter of the content comes from .org.au domains, the extent to which this represents Indigenous organisations cannot be determined without further analysis of that content.

The amount of data from government domains varies considerably depending on the specific data set and is evidently influenced by both the nature of the event or activism and the period of collection. Data collected in earlier periods being lesser in amount perhaps skews representation in favour of government sources.

The lack of structured or controlled metadata relating to all the resources in the data sets make it very difficult to determine the voices represented. The raw data sets include URLs, titles and content from metadata fields created by the original publishers. This could be investigated further to identify more specific voices to some extent; but the data is unlikely to be consistent – certainly not consistently applied – nor substantive in identifying organisations, people or geographic location.

Social media

The social media archive currently has 1,000 queries active for collecting out of a total of 1,764 (query = baseline hashtags and keywords). This doesn't include subscriptions and follows (approx. 800 currently). Of the active queries, 83 relate to Indigenous collecting, or about 8%. These collections are particularly valuable in documenting how quickly views and ideas change.

A search was run across social media collections in 2020 relating to the four nominated national campaigns. This includes posts from Twitter, Facebook and Instagram. It is based on post content rather than specific hashtags.

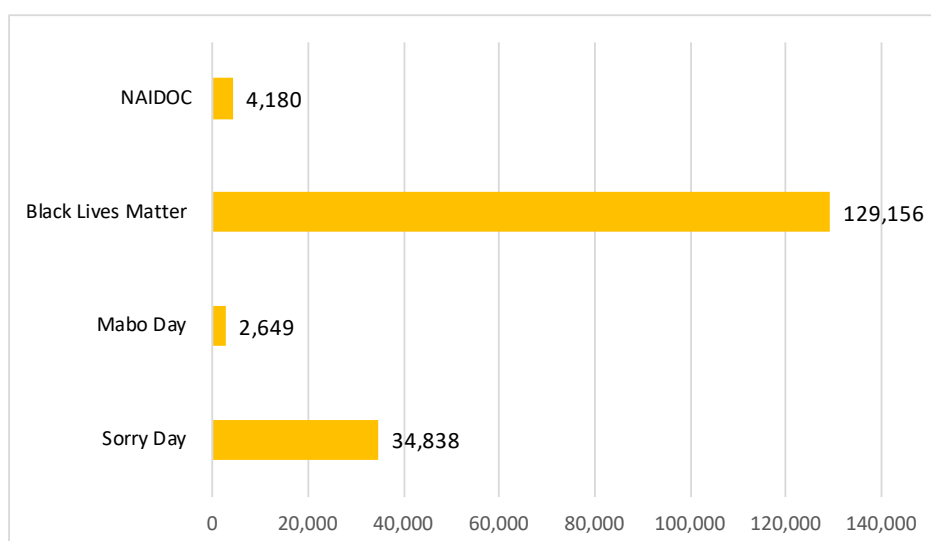FIGURE 22: NUMBER OF SOCIAL MEDIA POSTS COLLECTED BY SLNSW IN 2020 RELATING TO THE FOUR NATIONAL CAMPAIGNS.

The charts below present data for Black Lives Matter in June 2020 in various ways that can shed light on authorship and content. This data has been generated for each of the four campaigns in 2020.
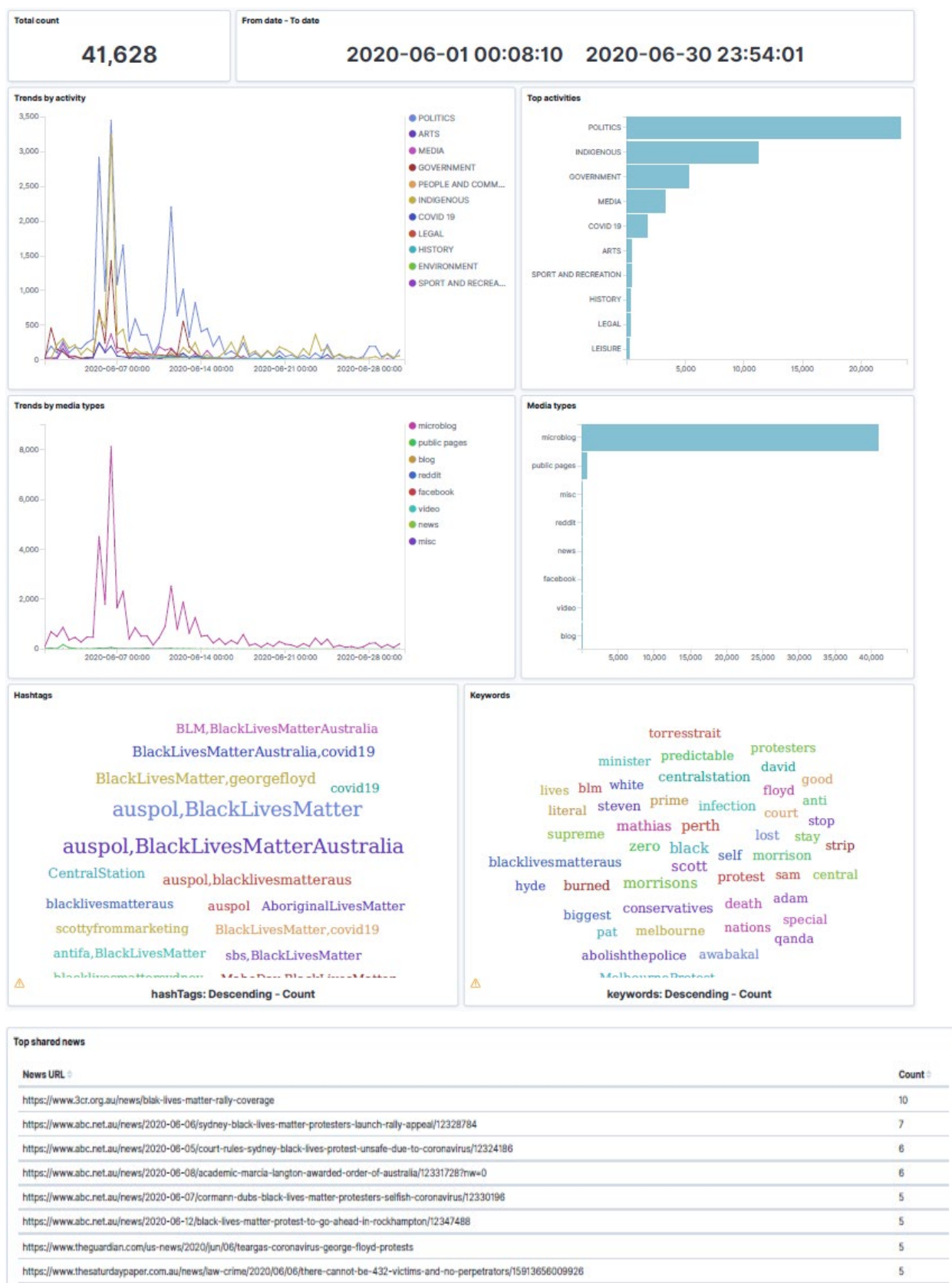

FIGURE 23: CHARTS AND WORD CLOUDS GENERATED FOR SOCIAL MEDIA ON BLACK LIVES MATTER, 2020.

Data for the Black Lives Matter sample shows social media content is coming from microblogs (i.e. Twitter) and public pages. News is being shared primarily from major media outlets (ABC, *Guardian*, *Saturday Paper*), although community radio takes the top spot here. Top authors and top mentioned authors are listed in the tables below.

Top authors	Top mentioned authors
Jeffskully [account suspended]	ScottMorrisonMP
Triplejay58	nswpolice
NACCHOAustralia	ellymelly
MigrantPartyAU [account suspended]	EddyJokovich
AusHandmaid [account suspended]	GladysB
PeterWMurphy1	LittleBertie01
AUSBATTERYGRAD1	adamajacoby
PeterPyke	ShoebridgeMLC

TABLE 5: TOP TEN AUTHORS AND MENTIONED AUTHORS IN BLACK LIVES MATTER SOCIAL MEDIA CONTENT COLLECTED IN JUNE 2020. GREY HIGHLIGHT INDICATES THOSE IDENTIFYING AS INDIGENOUS OR PUBLICLY SUPPORTED BY RECOGNISED INDIGENOUS LEADERS.

Of the authors above, just one is Indigenous-identified. A review of similar data across the four campaigns revealed that posts from political or religious groups were dominant, with few Indigenous organisations represented or readily identifiable despite the volume of Indigenous-related content.

While the existence of the State Library of New South Wales' social media archive is relatively well known within the library sector, it is not widely accessed by NSLA libraries or well-known publicly as yet. This may change following recent release of an API for accessing the metadata of the collection. While posts themselves cannot be shared due to privacy, data visualisation tools have been running for several years using an emotion clock based on anonymised and aggregated data. The archive captures a broad range of voices, but content is not yet well reflected in catalogue records.

At present, there is a very limited amount of social media collected as part of the Australian Web Archive through the PANDORA selective archiving process (aside from blogs) and this is mostly a small number of selected Twitter accounts. Moreover, this material is not distinguished from archived websites nor flagged as 'social media' per se. The SLNSW Social Media Archive is the only such collection by a national or state collecting institution.

Ephemera

The final set of collection material analysed in relation to the four nominated national campaigns was ephemera. Posters, pamphlets, banners and so on can be collected spontaneously and without the kind of official authorship that published collections demand. It was felt that ephemera collections were well suited for the documentation of Indigenous activism (indeed, any activism), reflecting community and Indigenous voices.

The findings from this exercise, however, revealed that ephemera collections were far smaller than anticipated.

	National Sorry Day	NAIDOC week	Mabo day	Black Lives Matter
SLWA	14 items	19 items, 1 box	0 items	0 items
SLNSW	2 boxes	2 boxes, 10 items	2 boxes	Not yet listed
SLQ	30 items (1 box)	446 items (5 boxes)	10 items	0 items
NLA	52 items (1 box), 1 poster	78 items (2 boxes), 29 posters	1 item, 2 posters	1 item, 1 poster

TABLE 6: EPHEMERA COLLECTIONS RELATING TO THE FOUR NOMINATED NATIONAL INDIGENOUS CAMPAIGNS.

Discoverability of this content was also less than ideal, with unpublished material usually discoverable at the item or box level only, and records unclear about the nature of the contents and if or how they could be accessed.

The potential was noted for ephemera to provide a unique lens on a community event, reflecting social views and messages from a range of participants in 'real time'. While web archiving activity will be catching the digital equivalent of some contemporary ephemera, it was pointed out that web and social media platforms capture material only from those with digital access and digital skills, while physical ephemera may be more inclusive of older communities, for example, or regional groups.

In the sample analysed, project members noted duplicate invitations, postcards and newsletters, again primarily reflecting government messages or propaganda rather than Indigenous voices. Most of this material has been donated or purchased by libraries. There were questions about what stories we are telling in the collections that we have, and whether libraries are in fact still collecting physical ephemera in any meaningful or strategic way. If not, should libraries reinvigorate this part of our collecting activity, and how might we do so with a view to capturing a far greater diversity of voices?
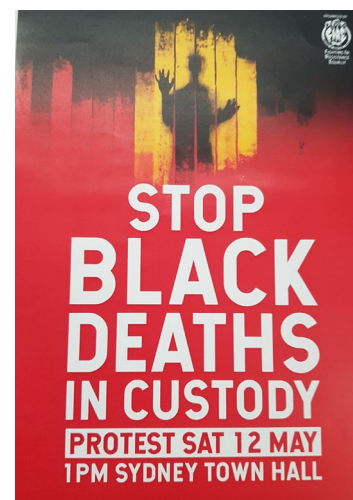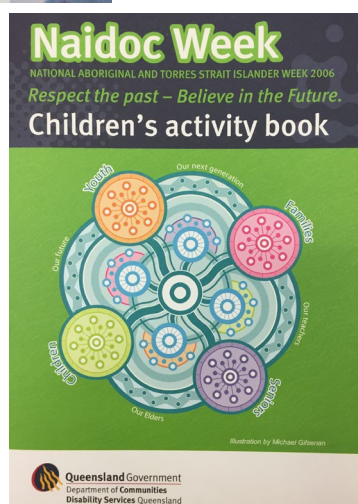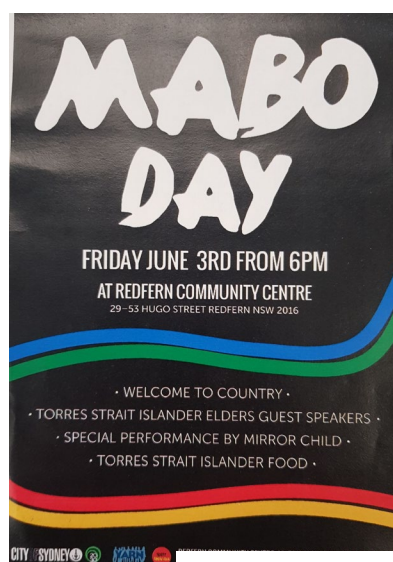
On this page: Examples of ephemera and ephemera boxes from State Library of Queensland and State Library of New South Wales

Conclusions

The contemporary Indigenous collections of NSLA libraries are vast. This audit alone – covering four libraries and a limited sample of content – includes over 18,000 publications; 17,500 websites; and 170,000 social media posts.

Commendable efforts have been made in the last two years particularly to include Indigenous languages and place names in publication records – with just under a third of the published collection including a language from the Austlang thesaurus and a similar number including a place name from the AIATSIS place names thesaurus.

The web archive dataset reveals a substantial amount of content that relates, in varying and not always apparent ways, to the Indigenous activism and events that were the focus of the second part of the audit. State Library of New South Wales' burgeoning social media collection provides another avenue for the collection of diverse voices, albeit constrained by the resources and spheres of reference of staff undertaking the collecting activity.

Metadata: the key to discovery

Despite the work that has gone into building rich and varied published Indigenous collections, results in this report show us that application of metadata, including from AIATSIS thesauri, is very inconsistent. Voices represented in the collection are heavily weighted to government over Indigenous and community, with only a sixth of the collection connected to identifiable Indigenous authorship based on the methodology chosen here.

Without consistent and high quality metadata, discoverability is impeded and opportunities diminished to curate collections from our vast archives in partnership with Indigenous authors or communities. Our capacity to draw meaningful statistics in order to inform selection and acquisition decisions, and to identify gaps, is also limited.

The metadata issues facing web collections relate more to scale and search functionality. Cataloguing of web archive content is very limited and has only been done for content collected for the PANDORA selective web archive (a small sub-set of around 8% of the entire Australian Web Archive corpus). Even archived websites that have been catalogued typically contain a great deal of content, including documents such as reports, submissions and papers, which will not have been analytically catalogued.

Access to relevant content in the AWA is primarily achieved through the full text search functionality of the 'websites' category in the Trove discovery service. Given the scale of this content (more than 14 billion documents) locating relevant content requires a skilled approach. Results from simple searching can be in the tens or even hundreds of thousands for many search terms. The AWA search index does support Solr advanced search syntax, but this is not a realistic option for most users. At present, AWA search results cannot be downloaded, limiting research to the active search session and providing no usable overview of search results.

Finally, this analysis has revealed an ad hoc approach to physical ephemera collections relating to Indigenous matters, and a weighting toward government materials, with general box-level descriptions reducing the likelihood that users will know of or access the items within.

Next steps

Most of these issues are known to collections staff in general terms, but being able to quantify them, and really testing the functionality of our own systems and the rigour of our own selection processes with this specific lens on Indigenous collections, has led to some practical recommendations from the project group below.

An analysis of current work in descriptive practices is provided at *Appendix A*, and a comparison between current policies and measures for Indigenous collections in participating libraries is provided at *Appendix D*.

Research questions relating to community engagement in the selection, acquisition, management, presentation and discoverability of original collections with Indigenous content will be explored in stage two of the project. This will be designed in July 2021 and will commence in the second half of the year.

Recommendations

A number of recommendations are put forward here for consideration by NSLA libraries, and by members of the NSLA Heads of Collections Advisory Group, for further action.

Published collections

- Use parent/child MARC tags to enhance discovery of related works
- Create a unique identifier (MARC) for Indigenous content
- Review Indigenous-identified datasets on a regular basis to inform acquisition decisions
- Establish a working group across NSLA libraries to:
  - Develop shared guidelines for description of Indigenous collections, including more uniform expressions of access or restriction, to improve consistency in material displayed in the ANBD
  - Draw on recent national and international work in descriptive standards and data governance for Indigenous collections, to make changes to practice – and particularly any recommendations from the current OCLC *Reimagining Descriptive Workflows* project (see *Appendix A*)
  - Collect unique material and devise strategies to increase the number of Indigenous authors and Indigenous voices represented
  - Map holdings by Austlang language location as a means of visualising the extent to which each language is represented in the NSLA collection
  - Build a stronger collaboration with AIATSIS to improve use of the AIATSIS thesauri

Web archiving

- Build upon and develop further the curated Pandora collections with an Indigenous focus by working with the communities involved
- Improve the discoverability of Pandora curated collections, e.g. by improving and using metadata associated with the collections; use services such as Dark & Stormy Archives
- Make use of existing PANDAS (database) metadata in Trove search and discovery, e.g. use API to add search result faceting; make use of and present Solr advance search functionality in user friendly ways
- Improve the usability of Trove web archive search and discovery results, e.g. by API searching and the ability to save (meta)data sets of search results

Social media

- Broaden the diversity of people contributing to contributing to social media collection, or providing advice, to ensure that collecting is not limited to a single perspective
- Promote use of the API to access metadata behind the social media collection
- Investigate the possibility of themed collections for public access

Ephemera

- Create finding aids or link data through hierarchy to allow for each item to be visible in the catalogue
- Develop a communications plan around the donation of physical and digital ephemera to increase public and regional awareness
- Include physical ephemera in collection strategies with an emphasis on diversity of voices represented
- Create an easy portal for the deposit of digital ephemera

Indigenous library staff and resourcing

- Address the prevalence of Indigenous-identified positions in NSLA libraries that are entry-level only, and provide secure, permanent, ongoing library and information management roles for First Nations staff
- Provide opportunities for Indigenous staff to be trained in cataloguing and description, so that First Nations staff such as Indigenous librarians and Indigenous library technicians can be responsible for describing Indigenous materials wherever possible

Appendices and attributions

Appendices

- A: Summary of recent work in descriptive standards and data governance
- B: Methodology for web archive
- C: Research questions and methodology for published collections
- D: Comparison of Indigenous collection policies and reporting measures in participating libraries

Project team

Michela Goodwin, NLA   Sean Volke, SLNSW

Paul Koerbin, NLA   Kerry-Ann Tape, SLNSW

Rebecca Bateman, NLA   Robyn Hamilton, SLQ

Nicole Lockwood, NLA   Debra Jones, SLWA

Ronald Briggs, SLNSW   Barbara Lemon, NSLA

- Data analysis (published items):   François Petitjean (with Michela Goodwin)
- Data analysis (web archives):   Paul Koerbin
- Data analysis (social media):   Sean Volke
- Descriptive standards review: Janice van de Velde
- Project management and final report: Barbara Lemon

The NSLA office thanks Kirsten Thorpe and Rose Barrowcliffe of the Indigenous Archives Collective for advice.

Datasets (available on request)

- List of PANDAS resources with Indigenous content
- List of AWA sites retrieved with search terms for National Sorry Day: 1998, 2008, 2018
- List of AWA sites retrieved with search terms for NAIDOC Week: 1996, 2006, 2016
- List of AWA sites retrieved with search terms for Black Lives Matter: 2013, 2020
- List of AWA sites retrieved with search terms for Mabo Day: 1996, 2006, 2016
- Social media collection profiles for National Sorry Day, NAIDOC Week, Black Lives Matter and Mabo Day in June 2020 [data from 2012 and 2017 also provided for NAIDOC Week]
- List of ephemera held by the four participating libraries relating to the four national campaigns
- Trove extract for published collections with Indigenous content x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of unique holdings x 4 (NLA, SLQ, SLWA, SLNSW)
- List of records including name of an Aboriginal Land Council x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of records including the name of an Aboriginal author-AustLit x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of records including an Indigenous language & aiatsisl tag x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of records including an Indigenous place name & aiatsisp tag x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of records including 'aiatsiss' tag x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of all record types represented x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of all bibliographic levels represented x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of physical versus digital records x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of all carrier types represented x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of records containing a URL in MARC 856 field x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of records deposited in NED x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of records with access conditions x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of records with cultural sensitivity messages x 5 (NLA, SLQ, SLWA, SLNSW, ALL)
- List of publisher names and number of publications associated with each
- Set of histograms showing distribution of collections over time, by publication date and digital format
- List of items with content in the MARC 500 field

Appendix A: Summary of recent current work in descriptive standards and data governance

The Australian Research Data Commons is a federally-funded national research infrastructure provider, which is providing funding for the Indigenous Data Network (IDN). The IDN, currently hosted by the University of Melbourne, aims to *help Indigenous communities in developing the technical capability and resources to enable them to manage their data for community advancement. By strengthening communities' agency in their data, the network empowers them to make informed decisions about their own development.* The first priority of the IDN is to establish a working relationship with AIATSIS and others before turning attention to issues that include, for example, identifying best practice of the data management lifecycle, and which organisations to work with that are currently active in this field. Examples given include the National Collaborative Research Infrastructure Strategy (NCRIS). [7] The IDN Roadmap Framework[8] identifies a key partnership role for libraries and archives in the restoration of data ownership by Indigenous peoples. The IDN adheres to the FAIR data management principles – findability, accessibility, interoperability and reuse of digital assets.

Local contexts is a web portal that provides an international system to support Indigenous communities to manage their intellectual and cultural property within the digital environment. Local contexts is built around two core tools: Traditional Knowledge (TK) and Biocultural (BC) Labels and Notices – these *visible digital identifiers* are intended to act as an extra-legal, educative metadata strategy. Most importantly they enable Indigenous communities to add *cultural and historical context and political authority to cultural heritage content in non-Indigenous digital archives, libraries, museums and other digital repositories globally as well as to their own digital heritage archives.*

Local contexts is also working with New Zealand's Te Kotahi Research Institute at the University of Waikato to develop TK licences, which are expected to be available in late 2022 for testing. The text for the TK Notice is applied at a metadata field level and cannot be changed, use of the Notice requires registration with the Local Contexts Hub, which will subsequently advise the community about its use. The Notices are intended to act *as place-holders on collections, data or in a sample field until a TK or a BC label is added by the community to replace it.* This is important as it can allow time for communities to develop and customise TK Labels. There are also two Cultural Institutions Notices – Open to Collaboration and Attribution Incomplete – designed to be implemented by institutions who are engaged in the process of collaboration and trust-building with communities.

Early in 2021 a joint OCLC-NSLA webinar discussed with library practitioners how institutions could improve descriptive practices based on a discussion around the CARE principles for Indigenous Data Governance, as well as the TK and BC Labels and Notices.[9] The CARE principles – Collective Benefit, Authority to Control, Responsibility and Ethics – have been developed by the Global Indigenous Data Alliance, an international network of Indigenous researchers, data practitioners advocated for Indigenous data sovereignty at the national and international level. The CARE principles complement the existing FAIR principles to facilitate open and sharing of scientific data, while emphasizing and supporting Indigenous innovation and self-determination.

Implementation of the CARE principles is also being supported by the Equity for Indigenous Research and Innovation Coordinating Hub (ENRICH) – an international network working to promote Indigenous protocols and practices to enable, among other things, Indigenous control of Indigenous data and Indigenous self-

---

[7] https://mspgh.unimelb.edu.au/__data/assets/pdf_file/0010/3055852/NIDN-2019-03-01-IndigenousDataNetwork-August2018.pdf
[8] https://mspgh.unimelb.edu.au/__data/assets/pdf_file/0011/3070487/NIDN-2019-03-08-Indigenous-Data-Network-Roadmap-190301A.pdf
[9] https://hangingtogether.org/?p=8901

determination. In addition to supporting practical implementation of the TK and BC Labels and Notices across institutions, and the CARE Principles ENRICH is developing training models and platforms to support data sovereignty agreements. ENRICH's Cultural Institutions Network predominantly reflects North American institutions, but does include the UK's Wellcome Collection, Canada's Simon Fraser University, the Australian Institute of Aboriginal and Torres Strait Islander Studies, as well as New Zealand's Auckland Museum, Auckland War Memorial Museum, and Ngā Taonga Sound and Film. ENRICH is also involved in an IEEE Working Group to establish standards and recommendations for metadata fields to support appropriate disclosure of original data information as well as long-term identification Indigenous data for future use.

In March 2021, OCLC announced it had received funding from the Mellon Foundation to undertake a Reimagine Descriptive Workflows project. This eight month project is working with the non-profit Shift Collective under the stewardship of an international advisory group, which includes representation from Australia and New Zealand.[10] OCLC notes that this project is *a significant landmark in repairing the bibliographic infrastructure that we all rely on [and] will allow OCLC and other stakeholders an opportunity to listen and gain a better understanding of the problem space and to create possible, scalable pathways to address harmful descriptive practices. The community agenda, as a published output, will be used to frame future global community conversations, inform OCLC product pilots, and to identify further research and learning opportunities. [11]*

AIATSIS has developed and maintains AUSTLANG as a thesaurus for Indigenous languages and peoples and information about Aboriginal and Torres Strait Islander languages. It uses alpha-numeric codes to function as persistent identifiers, followed by a 'string of changeable text' to accommodate changes to the name or spelling of a language according to community preference. These codes replace the single MARC language code 'aus' with alpha-numeric codes for hundreds of different language varieties: MARC language Source Codes list. The codes maintain an identity if a change is made to the spelling or the name.[12] AUSTLANG datasets (codes, reference names, authority headings and map locations) have been made available under a CC BY 4.0 licence since 2019. AUSTLANG has replaced the AIATSIS pathways thesaurus, but this is still used to describe items in the AIATSIS Collections.

New Zealand

In 2018, the Te Mana Raraunga, Māori Data Sovereignty Network, released a statement on Principles for Māori Data Sovereignty - authority, relationships, obligations, collective benefit, reciprocity, guardianship – which have similarities to the ATSILIRIN protocols. Te Mana Raraunga is committed to: asserting Māori rights and interests in relation to data, ensuring data for and about Māori can be safeguarded and protected, requiring the quality and integrity of Māori data and its collection, advocating for Māori involvement in the governance of data repositories, supporting the development of Māori data infrastructure and security systems, and supporting the development of sustainable Māori digital businesses and innovations.

The National Library of New Zealand has consistently participated in international programs and projects to support ICIP, some of which are noted above. The Library's descriptive cataloguing practices include use of Māori Subject Headings to incorporate the Māori worldview, standardised terms for Iwi and Hapū names,[13] however the Library notes that these shouldn't be seen as *authoritative beyond their use in libraries and archives.*

---

[10] Damien Webb, Manager Indigenous Engagement Branch SLNSW; Celia Joe-Olsen, Heritage Advice Coordinator, Alexander Turnbull Library, NLNZ. National Council Te Rōpū Whakahau. IFLA Indigenous Matters Standing Committee and Katrina Tamaira, Research Librarian Māori, Alexander Turnbull Library, National Library of New Zealand. Archifacts Editor: Journal of the Archives & Records Association of New Zealand Te Huinga Mahara.
[11] https://www.oclc.org/en/news/releases/2021/20210330-reimagine-descriptive-workflows-project.html
[12] https://collection.aiatsis.gov.au/austlang/about
[13] https://natlib.govt.nz/librarians/national-library-services/cataloguing

Appendix B: Methodology for web archive

Dr Paul Koerbin, NLA

Datasets were extracted for materials relating to the specified Indigenous events and activism from the Trove Australian Web Archive (AWA). The AWA is full text indexed and is publicly searchable through the Trove discovery service. However, it is not possible to download datasets of search results from the public discovery service. For this reason, the Solr Admin application was used to run advanced searches with results in CSV format which could then be download. These have been saved as Excel spreadsheets for each target event.

The AWA index is a full text index and as such does not consist of curated or structed metadata. Consequently, the search strategies used for each event were varied to address the nature and specificity of the search terms and the period searched. The objective in each case was to produce a relevant and substantial result, not every mention of the search terms.

While the search index is essentially free text, some specific fields could be searched to increase the relevance of the results, including the 'title', 'metadata' and 'h1' (i.e. page heading) fields. The searches were also able to be limited to year (or years) of collection (acquisition). A range of acquisition years was specified for earlier content (pre-2000) since web archiving procedures, technologies and activities were still being developed in this period and thus the was often a lag time between publication and collecting the material. From 2005 to 2010 annual domain harvests were conducted which were run in the later part of the calendar year; however from 2011 domain harvests have been conducted around March-April and for this reason searches for content after 2011 also included a range of years since the target year's data may not be collected until the following year's domain harvest.

The datasets include the fields: 'year' (year of acquisition), 'site' (the domain level URL), 'title' (page title), 'URL' (page URL) and 'metadata' (publisher created metadata, i.e. <meta> fields, where these exist). The spreadsheets have been left in the default sort order, which is the search results ranking order, but these can be reordered to bring together all content from a particular 'site', for example.

Some deduplication has been applied to the datasets – using the utility: awk '!x[$0]++' – to remove duplicate lines. This has not removed all duplication of content – for example where the same page/content has a slightly different URL, title or metadata – but does as least limit the amount of duplication.

The table below for each of the 15 datasets includes the Solr search string used; the list of fields in the delivered search result (the same for all datasets); the original number of search results; and the number of search results after deduplication.

1. National Sorry Day

1.1 1998

(1-1_1998_National_Sorry_Day.xlsx)

Search strategy	Field list	Original	Dedup'ed
(title:"sorry day" || title:"stolen generations" || title:"bringing them home" || title:"national apology" || title:"national day of healing") && 1998 year:[1998 1999]	year,site,title,url,metadata	160	91

1.2 2008

(1-2_2008_National_Sorry_Day.xlsx)

Search strategy	Field list	Original	Dedup'ed
(title:"sorry day" || title:"stolen generations" || title:"bringing them home" || title:"national apology" || title:"national day of healing") && 2008 year:2008	year,site,title,url,metadata	3129	3056

1.3 2018

(1-3_2018_National_Sorry_Day.xlsx)

Search strategy	Field list	Original	Dedup'ed
(title:"sorry day" || title:"stolen generations" || title:"bringing them home" || title:"national apology" || title:"national day of healing") && 2018 year:[2018 2019]	year,site,title,url,metadata	8509	5377

2. NAIDOC Week

2.1 1996

(2-1_1996_NAIDOC_Week.xlsx)

Search strategy	Field list	Original	Dedup'ed
("NAIDOC week" || "NADOC week" || "NAIDOC committee") 1996^4 year:[1996 1997]	year,site,title,url,metadata	66	46

2.2 2006

(2-2_2006_NAIDOC_Week.xlsx)

Search strategy	Field list	Original	Dedup'ed
(title:("NAIDOC week" || "NADOC week") || "NAIDOC committee") 2006^4 year:2006	year,site,title,url,metadata	839	759

2.3 2016

(2-3_2016_NAIDOC_Week.xlsx)

Search strategy	Field list	Original	Dedup'ed
title:("NAIDOC week" || "NADOC week" || "NAIDOC committee") 2016^4 year:[2016 2017]	year,site,title,url,metadata	4778	3580

3. Black Lives Matter

3.1 2013[2014]

(3-1_2013_Black_Lives_Matter.xlsx)

Search strategy	Field list	Original	Dedup'ed
"black lives matter" year:[2013 2014]	year,site,title,url,metadata	30	28

3.2 2020

(3-2_2020_Black_Lives_Matter.xlsx)

Search strategy	Field list	Original	Dedup'ed
title:"black lives matter" 2020^4 year:[2020 2021]	year,site,title,url,metadata	12213	4445

4. Mabo Day

4.1 1996

(4-1_1996_Mabo_Day.xlsx)

Search strategy	Field list	Original	Dedup'ed
("mabo day" || "mer island" || "eddie mabo") && 1996^4 year:[1996 1997]	year,site,title,url,metadata	74	52

4.2 2006

(4-2_2006_Mabo_Day.xlsx)

Search strategy	Field list	Original	Dedup'ed
(title:("mabo day" || "eddie mabo"~2 || "mer island") || metadata:"mabo day"|| h1:"mabo day") && 2006^4 year:[2006 2007]	year,site,title,url,metadata	44	42

4.3 2016

(4-3_2016_Mabo_Day.xlsx)

Search strategy	Field list	Original	Dedup'ed
(title:("mabo day" || "eddie mabo"~2 || "mer island") || metadata:"mabo day" || h1:"mabo day") && 2016^4 year:2016	year,site,title,url,metadata	360	217

The 15 datasets were individually analysed to break down the content by the principal 2nd level domains, i.e. org.au, edu.au, gov.au, net.au, com.au and asn.au. A count was also done to identify non.au content. The following command line utilities were used to extract and count the analysed content:

- grep '.org.au' file.csv | awk 'END {print NR}' (example of 2nd level domain analysis)
- grep -v '.au' file.csv | sed -n '1!p' | awk 'END {print NR}' (to count non .au content)

Note that the same process was applied to four local campaigns nominated by the participating libraries. Results for these are not included in this report. Local campaigns were:

- [National Library] ACT gazetted public holiday: Reconciliation Day 31st May
- [State Library of Queensland] Laura Quinkan Dance Festival [July 2021]
- [State Library of Western Australia] Mowanjum Festival [July]
- [State Library of New South Wales] NSW Koori Rugby League Knockout

Methodology for data extraction from the PANDAS (PANDORA Archive) database

The more substantial amount of Indigenous related content collected through PANDORA, and that which is provided in the provided spreadsheet, is based on subject listings not on collections. Subject listings cannot be browsed through the Trove discovery service but are accessible through the PANDORA public website (at the link given above). This includes the top-level listing 'Indigenous Australians' and 11 sub-categories including: Government Indigenous Policy; Indigenous Art; Indigenous Business & Commerce; Indigenous Culture; Indigenous Education; Indigenous Employment; Indigenous Health; Indigenous History; Indigenous Land Rights; Indigenous Languages; and, Indigenous Native Title.

A dataset of 3,108 titles consisting of PANDAS metadata was derived using Metabase in February 2021 for all listings under the subject 'Indigenous Australians' and the 11 Indigenous sub-categories. As some titles (archived websites or documents) will appear in more than one subject sub-category there will be some duplication. Generally, a title appearing under the broad top-level subject 'Indigenous Australians' was not considered specific enough to be listed in one or more of the sub-categories.

The dataset in the spreadsheet titled 'ICA_PANDORA_titles_23_Feb_2021.xlsx' includes the following fields for each identified resource:

- PI (PANDAS database persistent identifier number).
- Link to the archived 'title entry page' (this will direct to Trove) for access to the archived copies of the title.
- Date that the title was registered in the PANDAS system.
- The seed URL used to collect the title.
- Short display title.
- Subject listing name.
- Organisation (PANDORA partner) responsible for the archiving the title.
- First harvest date.
- Last harvest date.
- Gather schedule (if current; otherwise noted as 'none').
- PANDAS status (with 'ceased' meaning no longer being regularly harvested).

Appendix C: Methodology for published collections

MARC field	Description
000	Leader – selected codes *Coded data for bib format codes (06/07 positions) separated into a column and macro added for readable text*
001	Control Number – Australian National Bibliographic Database (ANBD) system number
008	008 Fixed Field – full string *Coded data separated in columns: e.g. date 1 & 2; place of publication code; govt publication; language code*
020 022	ISBN – multiple columns (14) ISSN
035	System Control Number – multiple columns (5) *Includes OCLC numbers, NED (National edeposit) system numbers*
041	Language Code – multiple columns (3)
043	Geographic Area Code
100 110 111	Main Entry - Personal Name Main Entry - Corporate Name Main Entry - Meeting Name
245	Title (e.g. Title : Subtitle / Statement of Responsibility)
260, 264	Imprint (e.g. Place of publication : Publisher, Year of publication) 260 – multiple columns (2) 264 – multiple columns (5)
300	Description
336, 337, 338	Content, Media, Carrier RDA fields 336 Content – multiple columns (4) 337 Media 338 Carrier – multiple columns (2)
500	General Note – multiple columns (6)
506	Restrictions on Access Note – multiple columns (3)
542	Information Relating to Copyright Status Note
546	Language Note
600	Subject Added Entry-Personal Name – multiple columns (35)
610	Subject Added Entry-Corporate Name – multiple columns (36)
611	Subject Added Entry-Meeting Name – multiple columns (2)
630	Subject Added Entry-Uniform Title
648	Subject Added Entry-Chronological Term
650	Subject Added Entry-Topical Term – multiple columns (49)
651	Subject Added Entry-Geographic Name – multiple columns (22)
653	Index Term-Uncontrolled – multiple columns (17) Australian ; Other subject terms
655	Index Term-Genre/Form – multiple columns (8)
656	Index Term-Occupation – multiple columns (2)
658	Index Term-Curriculum Objective – multiple columns (2)
700	Added Entry-Personal Name – multiple columns (4)
856	Electronic Location and Access – multiple columns (2)
984	Library holdings – multiple columns (5)
Holdings summary	4 NSLA participating libraries – multiple columns (5)

National data analysis: published collections

Apply a core set of questions to the master dataset and to each local dataset for comparative purposes based on MARC fields (see MARC fields summary above).
Use searches across the relevant fields – and external databases where relevant and listed in the table below – to provide the following:

Analysis task	Relevant MARC fields	External databases	Comments from analyst
Unique holdings			
List and number of unique holdings contained in the dataset.	984 holdings		
Voices represented			
List and number of publications that include the name of an Aboriginal Land Council and/or an Aboriginal author registered with AustLit.	100, 110, 111 Main Entry Name 700 Added Entry Personal Name (we do not have 710or 711) 600 - Subject Added Entry - Personal Name 610 - Subject Added Entry - Corporate Name 260 Publication, Distribution, etc. (Imprint) 264 Production, Publication, Distribution, Manufacture, and Copyright Notice 542 Information Relating to Copyright Status Note	AustLit list of authors List of Aboriginal Land Councils in Australia, Creative Spirit	List provided separately – files starting with Q2-1. The last column provides the Aboriginal Land name that was matched against the record. Note on methodology. Aboriginal Land Council rarely appear with the full name described in the list of Aboriginal Land Council. For instance, "Walgett Local Aboriginal Land Council" doesn't appear in the dataset, but some titles include "Walgett Aboriginal Medical Service Co-operative". I thus simplified the list of Aboriginal Land Councils. For instance, we looked for the appearance of the words 'Walgett', or 'Winbar' in the given MARC fields ('100-MARC','110-MARC','111- MARC','600-MARC','610-MARC','260-MARC','264-MARC','542-MARC','700-MARC'). As it wasn't clear which MARC field was for Aboriginal Land Council or for Aboriginal authors, I used the same for both. Numerous names in the AusLit list where listing only a first name and many were very common Anglo first names. To avoid too many false positives, I filtered the AusLit list down to only include non-anglo first names when only a first name was provided. To do this, I removed the top-1000 names registered in Scotland from 1880-2008. All other names were kept.

Languages			
List and number of publications that have at least one Austlang Indigenous language code included in their metadata.	650 - Subject Added Entry - Geographic Name with Source code: aiatsisl See also 546 Language Note 041 Language Code 008 Fixed Field: Language	Austlang list of Indigenous languages Austlang dataset	I used both 'language name' and 'language synonym' columns in the AustLang dataset. Both had often lists that I prefiltered to remove common terms that might not designate a language so as to avoid false positives (eg 'southern', 'north', 'mount'). 11,504 terms were kept to designate languages. The list is provided for reference in 'list-language-terms.xlsx'. I used the following MARC fields: '008-MARC','041-MARC','546-MARC', as well as all '650- MARC' fields.
Place names			
List and number of publications that have at least one Indigenous place name included in their metadata.	651 - Subject Added Entry - Geographic Name 651 - Subject Added Entry - Geographic Name with Source code: aiatsisp See also 043 Geographic Area Code	AIATSIS thesaurus of Indigenous place names	We have had difficulties (with Barbara and Michela) to extract terms from the aiatsisp thesaurus that refer to Indigenous places, because the thesaurus also includes the main regions. So we decided upon a two-pronged approach: 1. Creating a reduced list of terms from the thesaurus, based on the most used terms, and removing terms that are not linked to Indigenous place names (list of terms removed in methodology). 2. Finding the records that hold the term 'aiatsisp' in any of the 651MARC fields. We extracted all terms from the thesaurus (having removed map information) and removed the following terms: ['asia','africa','europe','japan','australia','china','india','new zealand','south east south australia','united states', 'germany','malaysia', 'indonesia','france','england','south africa', 'turkey','fiji','united kingdom', 'taiwan','timor','vietnam','zimbabwe','italy','palestine','brazil','sweden','greece','afghanistan','kenya','cuba','victoria','new south wales','western australia','tasmania', 'northern territory','south australia', 'queensland', 'australian capital territory', 'canada','great britain','papua new guinea','sydney','sydney harbour', 'melbourne', 'darwin', 'canberra', 'adelaide', 'hobart', 'wollongong','perth','brisbane','newcastle','london, england','melbourne map area', 'sydney map area','darwin map area','canberra map area','adelaide map area','hobart map area','wollongong map area', 'point', 'york', 'bell', 'peninsula', 'snowy', 'mer','yarra','beach','three','rocks']
Formats			
List and number of formats represented in the dataset.	Leader 06 and 07 positions for bib format		The MARC-300 field doesn't always contain the information about whether the record is digital or physical. We consider a record to be digital

	e.g. monographs, serials, notated music etc. 300 Physical Description – to identify specific physical and digital formats 338 Carrier Type – to identify online resources and other carriers. See also List of carrier types 856 Electronic Location and Access – to identify digitised collection items, NED publications (born digital) and purchased / subscription digital content (e.g. from EBSCO)		if the MARC-300 field contains any of the following list of terms: ['online','digital','cd-rom', 'sound','min','videodisc', 'videocassette', 'dvd', 'cd','disc', 'vhs','hdcam']. Note that a few records had invalid format (for carriers) but I didn't modify them. The MARC-856 field mostly contains either nothing or a URL. I extracted here the records for which that field contained the term 'http'.
Availability			
List and number of items with access conditions/restrictions.	506 Restrictions on Access Note Search for: "not for inter-library loan" "not for inter library loan" "not for interlibrary loan" "not for loan" "not for general viewing" "restrictions" "restricted" "available onsite" "access to be determined" "access closed"		We extracted the records that would have their MARC-506 field contain any of the following: 'not for inter-library loan','not for inter library loan','not for interlibrary loan','not for loan', 'not for general viewing','restrictions','restricted','available onsite','access to be determined','access closed'.
List and number of items deposited with the National edeposit service (NED).	035 System Control Number – contains NED system numbers, starts		Extracted all items that had the term 'NED' in MARC-035.

	with (AU-CaNED) 984 holdings – to identify NED holdings e.g. NED:NSL National edeposit [for State Library of NSW]		
List and number of items containing cultural sensitivity messages.	500 General Note Search for "sacred", "ceremonial" or "deceased"		We extracted the records that would have their MARC-500 field contain the words 'sacred', 'ceremonial' or 'deceased'.
Publishers			
Listing of publisher names and number of publications associated with each (to identify key publishers). [Note 264 field includes location, publisher, year – where publisher is the second term, after the colon]	260 Publication, Distribution, etc. (Imprint) 264 Production, Publication, Distribution, Manufacture, and Copyright Notice Publications without an 020 ISBN / 022 ISSN may indicate self-published / small press titles (we did not request 024 ISMN)		We extracted the names in MARC-264 $b (after ':' and before ','). We compared the fields without leading and trailing white spaces and in lower case. If no colon was present, then we considered that no publisher was present in the field. Formatting issues exist in this field.
Collecting trends			
Analysis to indicate if there has been an increase in digital collecting over the past five years.	008 Fixed Field: Date 1; Date 2; Place of publication 856 Electronic Location and Access – to identify digitised collection items, NED publications (born digital) and purchased / subscription digital content 984 holdings		For histograms showing distribution of collections over time: I use Date 1 in the 008 field by default and Date 2 if and only if there is a value in Date 2 and it precedes Date 1.

Requests from participating libraries

Requesting library	Task	Results/comments
SLWA	List and number of publications per region (WA only)	I looked for the provided regions in all the 264-MARC and 650-MARC fields. Goldfields-Esperance: 0 records Goldfields - Esperance: 0 records Peel: 3 records Kimberley: 176 records Great Southern: 0 records Mid West: 1 records Pilbara: 48 records Gascoyne: 13 records Wheatbelt: 5 records Perth: 315 records South West: 12 records
SLQ	List of items published by identified Indigenous publishers in QLD: • University of Queensland Press • State Library of Queensland • QAGOMA • Griffith University • Keeaira Press • John Wiley & Sons Australia, Ltd	Can be taken directly from publisher count spreadsheet in national analysis.
ALL	List and number of items published outside capital cities of Australia	**Could not be achieved.** The vast majority of items had a capital city listed as place of publication even if the publication derived from a regional area.
NLA	List of photographic material with titles including generic descriptors like 'man, men, woman, women, child, children, family, families, native, natives, black, blacks'.	**Could not be achieved.** Imperfect methodology to ascertain that a record is or contains photographic material. Even based on an imperfect list, these generic descriptors could be located in multiple locations. Data extracted from Trove did not cover all relevant fields. Agreed to undertake in-house at NLA.
NLA	List of items actually published between 2015-2020 (as against records enhanced in that period).	Unfortunately, there is no way to distinguish between when the library received the item into the collection rather than when the record was updated. The item may have been published in 1980 but purchased it in 2015.

Appendix D: Comparison of Indigenous collection policies and reporting measures

Collections[14]				
State Library of Queensland[15]	State Library of New South Wales[16]	National Library of Australia[17]	AIATSIS[18] key performance indicators	Potential NSLA performance indicators[19]
Actively collect material by Aboriginal and Torres Strait Islander people to ensure collection reflects diversity of Aboriginal and Torres Strait Islander experiences	Materials, published and unpublished, from Indigenous and non-Indigenous organisations, individuals, communities and events that document, articulate and animate Indigenous life, past and present in NSW.	Acquire published and unpublished material relating to all issues concerning First Nations Australians. Work in collaboration with AIATSIS to support the acquisition and preservation of Indigenous language and related collections.	KPIs for the strategy are: Setting up mechanisms to measure the impact of: • Engaging with community on collecting, describing and making material available; and • Providing access to collection materials and providing digital copies Building a representative collection	# and/or % collections items by or about Aboriginal and Torres Strait Islander peoples acquired through all acquisition methods and collection formats – not currently collected. # and/or (% increase) in collection building by language, community group, geographic regions – not currently collected.
Gain a clearer understanding of the breadth and representation of Aboriginal and Torres Strait Islander experiences to	Collect material, including commissions, that illustrates individual contributions…as well as material representative of everyday/lived experience	Develop the breadth and depth of the Library's First Nations Australian collections Ensure that First Nations Australians perspectives and	Reporting measures: • Efforts undertaken to improve representative collection building by:	NSLA annual audit q. 19, 21

---

[14] Headings 'Collections', 'Description' etc. adopted from SLQ Collections Commitments statement

[15] State Library of Queensland, Aboriginal and Torres Strait Islander Collections Commitments

[16] State Library of New South Wales, Indigenous Collecting Strategy

[17] National Library of Australia, First Nations Australians Collecting Principles

[18] AIATSIS, Collection Development Strategy 2017-2021 Performance Indicators – quantitative and qualitative indicators 'will be used to support progress and impact' p. 17
https://aiatsis.gov.au/sites/default/files/2020-09/collectiondevelopmentstrategy20172021.pdf

[19] These are largely qualitative, column identifies alignment with 2021 NSLA annual survey collection related questions, and alignment to ATSIRILIN protocols

| inform strategic and targeted collection building.

Work with Aboriginal and Torres Strait Islander communities to identify, contextualize and prioritise materials to collect | Consult with AIATSIS to support and assist in collection development.

Collection strategy will follow ASTILIRIN protocols, especially Protocol 2, Content and Perspectives.[20] | cultural requirements drive the collection for First Nations material

Actively engage with First Nations Australian individuals, communities and organisations to help build relevant and well-understood First Nations Australian collections.

Anticipate the needs of future generations of First Nations Australians.

Prioritise the collection of documentary heritage created by and relating to First Nations Australians, for the benefit of all Australians.

Build a collection that is beneficial and relevant to First Nations Australians. | language; community groups; geographic regions

• Number of collections for which an Aboriginal and Torres Strait community was engaged at some point in selecting, advising on access provisions or enhancing descriptions of materials in collections.

• Building a collection of significant Aboriginal and Torres Strait Islander material

• Percentage of offers accepted by significance ranking of materials by Aboriginal and Torres Strait Islander individuals and organisations;

• Recording both acceptance and rejection of offers and justification for decision-making. | Aligns with ATSILIRN Protocol 1, 2, 3, 10 |

---

[20] ATSILIRIN Protocol 2, Content and Perspectives: 2.1 Consult with an appropriate and ongoing manner with Aboriginal and Torres Strait Islander communities in regard to the development and management of the collections; 2.2 Seek to balance collections by acquiring material by and about Aboriginal and Torres Strait Islander peoples; 2.3 In the case of government archives, consult through the relevant government agency. Agencies should be advised of Aboriginal and Torres Strait Islander content of materials and appropriate access policies; 2.4 Promote the existence and availability of collections and provide clients with an explanation for any conditions governing access; 2.5 Facilitate the development of Aboriginal and Torres Strait Islander knowledge centres.

Description				
State Library of Queensland	State Library of New South Wales	National Library of Australia	AIATSIS performance measures	Potential NSLA performance measures
Retain clear records that demonstrate the context of collections and decisions made in relation to them. Use national Aboriginal and Torres Strait Islander description tools and classification systems which describe items by their geographic, language and cultural identifiers Provide opportunities for Aboriginal and Torres Strait Islander people to describe and annotate material that relates to themselves and their constituents.		Ensure that culturally appropriate description and access conditions are applied to First Nations Australians material at the time of collection and establish principles to guide this process.	Mechanisms to measure the impact of: engaging with community on collecting, describing and making materials available	NSLA annual audit q.23-25 Aligns with ATSILIRN Protocol 5, Description and classification processes

Access and use

State Library of Queensland	State Library of New South Wales	National Library of Australia	AIATSIS	Potential NSLA Performance Measure
Make Aboriginal and Torres Strait Islander collections as broadly accessible as possible. Promote collections to increase understanding of Aboriginal and Torres Strait Islander peoples' diverse experiences, history and cultural heritage. Adopt a tiered approach to assessing collections and making them accessible based on collection knowledge, context, community input and risk. Respond to requests to remove material from public access, utilising established procedures. Provide context and appropriate cultural care messages for potentially offensive material made publically accessible. Work with communities to identify secret, sacred and offensive material.		Ensure that culturally appropriate description and access conditions are applied to First Nations Australians material at the time of collection and establish principles to guide this process. Ensure that First Nations Australian knowledge, stories and voices in its collections are honoured, respected and made accessible in culturally appropriate ways. Establish and maintain the cultural safety of First Nations Australians material in its collections and of First Nations Australians accessing it.	Mechanisms to measure the impact of: engaging with community on collecting, describing and making materials available	NSLA annual audit q.20-22, 26 Aligns with ATSILIRN Protocols 3-12 NSLA annual audit q.22 Aligns with ATSILIRN Protocol *9*, Developing Professional Practice

Custodianship

State Library of Queensland	State Library of New South Wales	National Library of Australia	AIATSIS performance measure	Potential NSLA performance measure
Respond cooperatively to requests from Aboriginal and Torres Strait Islander communities for copies of material and information of relevance. Respond to requests for repatriation of original material to Aboriginal and Torres Strait Islander communities. Assist Aboriginal and Torres Strait Islander communities in planning, developing and managing keeping places for repatriated materials.	[ATSILIRIN Protocol 11]	Enable First Nations Australians to connect to information important to their lives, identities and cultural connections.		NSLA annual audit q.27 Aligns with ATSILIRN Protocol 11, 6

Capability

State Library of Queensland	State Library of New South Wales	National Library of Australia	AIATSIS	ATSILIRIN	Potential NSLA performance measures
Empower communities to access and manage their own cultural heritage at a local level. Ensure high levels of cultural competency and collection knowledge among State Library staff.					NSLA annual audit q.5 Aligns with ATSILIRN protocols 1, 2, 3