



Collections Unlocked: Open Data Project

Shared Leadership 2023

Authored by Kate McHugh, Karmel Savahl, Patrick Jovaras,
Christopher Nye and Jenny McKnight as part of the State Library of
Victoria & Public Libraries Shared Leadership Program 2023.
Sponsored by Victoria Whelan, Coordinator Libraries, Latrobe City.

Contents

Executive Summary	5
Project overview	5
Key findings	5
Recommendations	6
Background	7
The library landscape	7
Open data concepts	7
What is open data?	7
Open data in the library environment	8
Why open data?	9
Examples of open datasets	9
Use cases	12
Areas of concern	14
Privacy principles and legislation	14
Privacy and the Australian law	14
De-identification explained	15
Risk assessment	16
Data hosting and security	16
Evaluation	18
Stakeholder consultation	18
Methodology	18
Stakeholder theme 1: how can I protect myself from financial and reputational risk?	18
Stakeholder theme 2: how will the data be used and by whom?	19
Stakeholder theme 3: how will this be resourced, both in staffing and funding	19
Sample data set and application	21
Library services surveyed	21
Data collected - loans report	21
Data collected - holdings report	23
Data collected - size	24
Discussion	24
Challenges	24
Data analysis examples	25
Recommendations	29

References	30
Appendix 1: Stakeholder engagement mapping	32
Questionnaire	33
Stakeholder engagement interviews	33

Executive Summary

Project overview

The following report proposes the collation and sharing of key collections data across all Victorian public library services. It is proposed that this data be shared in a fully open format that can be readily accessed within libraries, as well as by interested stakeholders and the broader community.

The key loan data, which is proposed to be shared, is detailed on pages 21-24, along with example use cases (p. 12) and sample data analyses and reporting (pp. 25-28). Making Victorian public library service collections data openly available is associated with a number of potential benefits:

- Encouraging research and sharing insights to help drive collaboration, advocacy, and service innovation.
- Improving evidence-based decision-making.
- Meeting federal, state and local government requirements to make data open as a key pillar of open and transparent government.

A further discussion of the benefits of open data, examples of open datasets, and an overview of the concepts associated with open data, can be found on pages 7-9 of this report. To demonstrate the feasibility of this proposal and provide a 'proof of concept', a sample dataset drawing collections data from six Victorian public library services has been developed. These services were selected to represent a cross-section of library services in Victoria and also cover multiple library management systems (LMS) to provide insight into the merging of collections data from different systems (pp. 24-25). Interviews were also undertaken with key stakeholders from across the Victorian public library sector to gauge interest and identify areas of concern. Findings from this consultation can be found on pages 18-20.

Key findings

Following an analysis of the policy landscape, consultation with key stakeholders across the library sector, and technical scoping with a sample set of collections data from six Victorian public library services, the following key findings were identified:

- The creation of a collections open dataset would provide valuable insights for library staff working in the areas of collections development, management and promotion and would lift evidence-based practice, collaboration and service innovation.
- Whilst there is broad support for the creation of a consolidated open data reporting platform or dataset, stakeholder concerns around privacy, costs, scope, and potential for data misuse need to be addressed to progress this proposal further.
- Outcomes resulting from sharing collections data with broader stakeholders and the general public are unknown, however this lack of predetermination is also a necessary precursor to the innovation associated with open datasets.
- The process of combining data from multiple sources into a clean and usable dataset is not without its challenges, but certainly feasible. These challenges, the way data is

described and reports formatted, would be greatest during the implementation phase and extra support may be needed for some library services to develop reports to streamline the data collection process, through automated reports or API functions.

Recommendations

It is recommended that:

- The collation and publication of an open data set of Victorian library service loan and collections data is both feasible and necessary for the growth and development of the sector. A move towards open data as standard will demonstrate the sector's commitment to transparency and democratic access to information.
- Further education and advocacy work is undertaken across the Victorian public library sector, allowing staff to further familiarise themselves with the concept and the potential benefits. Recommended methods of advocacy include:
 - The distribution of this feasibility study amongst members of the ICT SIG, accompanied by a presentation from the project team on the findings
 - Agreement with the Public Libraries Victoria Network (PLV) to create a pilot test system, with a dashboard interface, that can be accessed by public library staff to grow comfort with the concept and foster confidence in evidence-based decision making
 - Recruitment of interested library services to volunteer their own data, i.e. through a .csv file to enable this pilot

Background

The library landscape

The landscape in public libraries has dramatically changed and evolved over time. Previously, information and data for decision making could be garnered locally and generated in-house by specialist librarians or teams via the LMS and via anecdotal understanding of the community. Collection development was predominantly undertaken by the Collections Manager via physical selection of items, having a detailed understanding of the communities reading preferences. In the past this information and data may have been channelled through by librarians who knew their customers' preferences and their reading history intricately and could influence selection. Collections could be easily curated to meet the community needs and wants.

With the introduction of technology such as RFID and evolving fast paced nature of libraries, the ability to garner data and information has changed. Technological advances such as self checkout via kiosks and return of items via book sorters have meant that there is less physical interaction between library staff and customers; thus, the ability to determine by whom items are being borrowed is limited. Whereas once Collections Managers were able to physically select items, there is now a shift to profile and standing order selection. Prior to these advancements, every item in the library would have been handled by the Libraries Technical Services Team or Cataloguing Librarian for cataloguing and end processing. This has been now largely outsourced to vendors or external agencies.

Libraries are a lot more fluid now, with an increase in patronage and an exponential increase in the breadth of services offered. Library staff are having to multi-skill to meet the needs and expectations of the community. Data and information were once anecdotal, held in the knowledge of the library service staff. Now, specialist knowledge is required to deliberately extract data from centralised repositories such as the LMS and other supporting collection development and management tools.

All of this data forms a rich vein of knowledge from which valuable insights can be mined in order to best meet community needs and expectations. Library service resources can be better utilised if the PLV were to share their expertise and access to data to ensure this understanding reaches across the state in a democratic way. Investment in open data across the Victorian library sector will prove longitudinally fruitful, allowing staff and stakeholders to harvest information, undertake analysis, and present useful findings to improve services and maximise ROIs.

Open data concepts

What is open data?

Open data is data that anyone can use: by design, it's accessible without restriction, machine readable, and able to be modified and reused. Open data can be used by anyone as they

wish, requiring only attribution to the original creator and a sharing, in turn, of any new knowledge or products stemming from the original dataset (Open Knowledge, n.d.).

To be considered 'open', data needs to be:

- published as close to its time of collection as possible
- organised in a format that allows for machine processing
- available in an open file format (CSV, XML, XBRL)
- made available in a permanent and stable format online under licence
- checked to ensure it does not include private, confidential or identifiable information, in order to comply with the *Privacy and Data Protection Act 2014* (Vic)
- Formatted to accepted data standards covering quality, metadata, exchange, access management (State Government of Victoria, 2020).

Whilst the idea of freely sharing data to benefit the advancement of knowledge is not a new idea, the concept of 'open data' officially dates to the 1990s, with the term first appearing in print in 1995 (Chignard, 2013). The open data movement can be considered as part of a broader movement towards 'openness' which has gained pace since the 1990s concurrently with the rise of the global Internet and World Wide Web. This movement encompasses areas including open access publishing, open government, open-source software, and open educational resources (Biswas-Diener & Jhangiani, 2017).

Since the establishment of the open data movement, governments and organisations globally have supported the move towards open data as a key aspect of open and transparent governance, and as a driver of innovation and opportunity. This includes the Australian Government, which has committed to making all 'non-sensitive data as open by default' and the State Government of Victoria, which has committed to making government data openly accessible, unless issues of law enforcement, public health, safety, or privacy apply (Australian Government, 2015; State Government of Victoria, 2021). Many Victorian local government organisations also have open data policies which encourage the sharing of open data (see, for example, City of Greater Geelong, 2023; City of Melbourne, 2023; City of Yarra, 2018).

Open data standards

To support government agencies and departments making data 'available for the public good', the State Government of Victoria has developed the DataVic Access Policy (State Government of Victoria, 2021a). The policy provides guidelines governing the release, licensing and management of datasets. Additional guidance has also been developed around information technology management and systems, and the publication of open datasets (State Government of Victoria, 2021b, State Government of Victoria, 2021c). These documents can, and should, also be adopted by local government agencies and specify format, protocols for data exchange, quality standards and metadata implementation.

Open data in the library environment

The objectives of the open data movement to ensure free access to, and use of, data would seem a good fit with the role and objectives of the public library to be the 'community's local source of communal information' (State Library of Victoria and Public Libraries Victoria Network,

2013). Libraries are in the business of providing free access to information, yet the sharing of metadata around collection usage (outside of dedicated repositories) is an area yet to be fully established. The International Federation of Library Associations and Institutions (IFLA) has recently called on libraries to improve the sharing of usage data (IFLA, 2021). According to IFLA, this will assist with learning and information exchange globally, help to bring a wider group of actors into conversations around libraries, encourage research, and share insights with less advantaged library services.

Why open data?

As mentioned above, both the Victorian and Australian Governments, along with an increasing number of local government agencies, are committed to ensuring government owned or generated data is made openly accessible (Australian Government, 2015; State Government of Victoria, 2021). This is due to the proven potential for open data to enable:

- Increased innovation, productivity, and economic growth
- Improved access to primary data sources by researchers, educators and the general public
- Increased efficiency and effectiveness of government-funded services
- Greater use of evidence-based decision-making
- Greater transparency of use of taxpayer and ratepayer funds
- The idea of government data as a resource which benefits citizens

Despite the wealth of data held by the government sector in Australia and the enormous opportunities enabled by recent improvements in data generation, analytical datasets and computing power, evidence suggests that the government sector is generally lagging in this space in Australia, with most innovation occurring in the private sector (Productivity Commission, 2017, p. 99).

Examples of open datasets

Despite some areas of government lagging in providing access to open data, there are some notable examples of improved government service-delivery enabled by open data in Australia. These include:

- The *My School* website, which supports transparency and accountability in school performance, and empowers families to compare different schooling options (Australian Curriculum and Reporting Authority, 2023)
- The Crime Statistics Vic website, providing data regarding offending, incidents and Victim reports which can be narrowed to local government area, postcode or suburb (State Government of Victoria, 2023).
- The City of Melbourne's *Urban Forest Visual* provides data on light, humidity, and level of canopy coverage in city environments to help government administrators make decisions regarding biodiversity and urban cooling (City of Melbourne, 2020).
- The Homes Victoria open datasets, which provide key statistics on the private rental market and housing affordability in Victoria (State Government of Victoria, n.d.).

Examples of Public Library Open Data initiatives

Public Libraries internationally are already independently creating and publishing open datasets. Notable examples of existing library sector-specific open data initiatives include the following library services based in the United States:

Seattle Public Library (USA)

<https://data.seattle.gov/Community/Checkouts-By-Title-Physical-Items-/5src-czff>

<https://data.seattle.gov/Community/Checkouts-by-Title/tmmm-ytt6>

Commencing in April 2005, the Seattle Public Library has produced monthly open datasets showing the number of checkouts by each of its holdings (including physical (book and other non-book lending materials) and electronic items) by title.

Table Preview

View Data

Create Visualization

ID	Checkou...	BibNumb...	ItemBarc...	ItemType	Collection	CallNum...	ItemTitle	Subjects	Checkout...
2005052315...	2005	1842225	0010035249...	acbk	namys	MYSTERY EL...	Loot a novel	Mystery ficti...	2005 May 23...
2005121417...	2005	1928264	0010037335...	jcbk	ncpic	E TABACK	Joseph had ...	Coats Fiction...	2005 Dec 14 ...
2005081113...	2005	1982511	0010039952...	jcvhs	ncvidnf	VHS J796.2 ...	know your yo...	Toys, Yo yos	2005 Aug 11 ...
2005101919...	2005	2026467	0010040985...	accd	nacd	CD 782.4216...	Tomorrows s...	Country musi...	2005 Oct 19 ...
2005122915...	2005	2174698	0010047696...	jcbk	ncpic	E KROSOCZ	Bubble bath ...	Baths Fiction...	2005 Dec 29 ...
2005100814...	2005	1602768	0010028318...	jcbk	ncpic	E BLACK	royal nap	Noise Fiction...	2005 Oct 08 ...
2005093010...	2005	2285195	0010053424...	accd	cacd	CD 782.4216...	real new Fall ...	Rock music ...	2005 Sep 30 ...
2005120517...	2005	2245955	0010048392...	jcbk	ncnf	J949.73 Or7...	Slovenia	Slovenia, Slo...	2005 Dec 05 ...
2005072215...	2005	770918	0010044828...	jcbk	ncpic	E HILL	Wheres Spot	Toy and mov...	2005 Jul 22 ...
2005050511...	2005	2288252	0010053488...	jcdvd	ncdvd	DVD J HOW ...	How do dino...	Violin Fiction...	2005 May 05...
2005060817...	2005	2129883	0010044896...	acdvd	nadvd	DVD GOOD TH	Good the ba...	Western film...	2005 Jun 08 ...
2005052716...	2005	1100696	0010039424...	acbk	nypb	FIC COONEY	face on the ...	Kidnapping F...	2005 May 27...

Figure 1. A snapshot of the Seattle Public Library open dataset log, which lists all physical item checkouts occurring from April 2005 to present day.

Timberland Regional Library Service (USA)

<https://www.trl.org/open-data>

Timberland Regional Library Services (TRL) “recognizes the rights of citizens, organizations, and local government to public access to library records and information” and provides open datasets relating to library service information, memberships and patronage, use of technology, collection circulation and financial information. TRL open datasets include raw data, in addition to a dashboard which enables data visualisation.

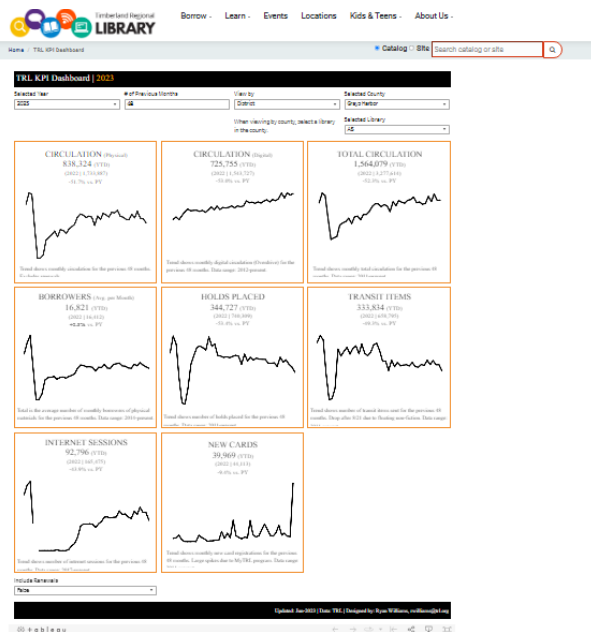


Figure 2. A snapshot of the TRL dashboard, displaying visualisation representation of the various TRL open datasets.

On a smaller scale there is an existing example of an Australian public library which publishes specific open datasets for a defined time period and specific purpose each month:

Brisbane City Council Library

[Library Checkouts - Datasets - Data | Brisbane City Council](#)

The Brisbane City Council library provides a monthly three-day snapshot (generally the 7th to 9th day of every month) of all of the resources borrowed from all of its branches. The open dataset shows line-by-line items checked out from Brisbane City Council branches during this defined period.

Library Checkouts — 7–9 Jun 2023 — CSV

Brisbane City Council Library checkouts borrowed at all branches on a single day.

Information includes: Title, Author, Call Number, Item Identifier, Item Type, Status, Language, Checkout Branch, Date Borrowed.

Data is a three-day snapshot of the month.

Resources	Additional Information																		
Library Checkouts — Heading acronym list — metadata — CSV	<table> <tr> <th>Field</th><th>Value</th></tr> <tr> <td>Date Published ⓘ</td><td>21/06/2023</td></tr> <tr> <td>Last Updated ⓘ</td><td>21/06/2023</td></tr> <tr> <td>Review Due ⓘ</td><td>20/07/2023</td></tr> <tr> <td>Update Frequency ⓘ</td><td>Monthly</td></tr> <tr> <td>Data Temporal Coverage ⓘ</td><td>07/06/2023 to 09/06/2023</td></tr> <tr> <td>Format ⓘ</td><td>ZIP</td></tr> <tr> <td>License ⓘ</td><td> Creative Commons Attribution 4.0</td></tr> <tr> <td>Attribution ⓘ</td><td>© Brisbane City Council 2023</td></tr> </table>	Field	Value	Date Published ⓘ	21/06/2023	Last Updated ⓘ	21/06/2023	Review Due ⓘ	20/07/2023	Update Frequency ⓘ	Monthly	Data Temporal Coverage ⓘ	07/06/2023 to 09/06/2023	Format ⓘ	ZIP	License ⓘ	Creative Commons Attribution 4.0	Attribution ⓘ	© Brisbane City Council 2023
Field	Value																		
Date Published ⓘ	21/06/2023																		
Last Updated ⓘ	21/06/2023																		
Review Due ⓘ	20/07/2023																		
Update Frequency ⓘ	Monthly																		
Data Temporal Coverage ⓘ	07/06/2023 to 09/06/2023																		
Format ⓘ	ZIP																		
License ⓘ	Creative Commons Attribution 4.0																		
Attribution ⓘ	© Brisbane City Council 2023																		
Library Checkouts — 7–9 Jan 2020 — CSV																			
Library Checkouts — 7–9 Feb 2020 — CSV																			
Library Checkouts — 7–9 Mar 2020 — CSV																			
Library Checkouts — 7–9 Apr 2020 — CSV																			
Library Checkouts — 7–9 May 2020 — CSV																			
Library Checkouts — 7–9 Jun 2020 — CSV																			
Library Checkouts — 7–9 Jul 2020 — CSV																			
Library Checkouts — 7–9 Aug 2020 — CSV																			
Library Checkouts — 7–9 Sep 2020 — CSV																			
Library Checkouts — 7–9 Oct 2020 — CSV																			

Figure 3. Links to the .csv files of the Brisbane City Council Library checkouts borrowed at all branches during a defined monthly period. Information includes: Title, Author, Call Number, Item Identifier, Item Type, Status, Language, Checkout Branch and Date Borrowed.

Use cases

Following a) the creation of the prototype sample open datasets (see **Sample datasets and application** section below), b) the consultation undertaken with a range of library sector and other relevant stakeholders, and c) the analysis and review of existing public library open data initiatives nationally and internationally, a range of potential use cases and benefits have been identified.

These include, inter alia, the following:

- **Funding, improvement to public library service access and advocacy**

From the open datasets, a heat map of library users across Victoria could be created based on postcode and number of loans.

This data could be used to assist the state government in identifying where funding should be directed based on high or low use of public libraries. In addition, it could be used at a statewide level to assist in the identification of areas where there is underrepresented or no public library presence, where public libraries require greater advocacy, or where there are potential risks of low literacy, accessibility issues, or disengaged communities.

The statewide snapshot and ability to compare similar local government or regional areas could provide ammunition for advocacy campaigns for individual library services and management within the sector and externally to funding bodies ie state and federal governments.

In general, the datasets could assist in raising all public library profiles. For example by providing data which could be adopted by journalists or media to generate stories -eg which suburbs have the largest number of library users, which book was the most borrowed etc.,

- **Collection development, management & promotion**

An obvious benefit from the creation of a statewide open dataset of public libraries holding loans by title is for collection management & procurement. This data could be used for comparison of library services by similar size, economic, geographical or social demographic.

Library services could use the data to collaboratively create evidenced-based standing order profiles, assist with patron-driven acquisition models or to identify collection gaps.

For example - This would be of great benefit where library services have limited funding or budgets, or limited local information that could assist in the introduction of a new niche collection (e.g. LOTE collection). The data could provide a detailed list of popular and high circulating titles and authors.

In order to assist in the promotion of collections, loan or cyclical usage patterns could be identified at a statewide level and used locally by individual library services for the promotion or even physical positioning of library collections.

For example, the sample datasets created for this project identified that items classified as 'Young Adult' were predominately borrowed by patrons aged over 35 years old, indicating that it would be prudent for libraries to physically position Young Adult collections adjacent to their Adult collections to increase borrowing.

- **Accurate, consistent and transparent reporting**

The access to statewide public library open datasets ensures the transparency and integrity of data that can be audited as it can be drilled down to a granular level. This would ensure accountability and consistent reporting by individual library services.

In addition, having centralised and open data repositories could place pressure on resource vendors to provide better reporting capabilities and sharing of data, which should be owned and accessible to the individual libraries subscribing to their services.

- **Use by authors, publishers and suppliers**

The obvious benefactors external to the library sector are the individual authors, publishers and suppliers of library resources.

Benefits for authors and publishers include:

- An understanding of where an author's work is most borrowed (for example the information could be utilised by the Melbourne writers festival to filter by author name to discover quickly where an author's work is most borrowed).
- Potential improvement to the Public Lending Rights Scheme which provides eligible Australian creators and publishers compensation for income that is lost through the public lending of their works through public library services.

- **Collaborative use by state and federal government**

Access to the open datasets could be used by the state and federal governments in conjunction with other existing data to determine whether the provision and access to library services and materials impacts on other social and economic factors.

For example, the state and federal governments could access the data in conjunction with other data repositories to analyse the following:

- Is there a correlation between high library usage and borrowing and literacy levels?
- What is the correlation between high library usage by youth to educational, vocational or employment outcomes?
- What is the correlation between high library usage and health and wellbeing and community engagement outcomes?

Areas of concern

Prior to the undertaking of stakeholder engagement, the project team immediately identified two areas of concern requiring investigation. The first area of concern relates to privacy of personal data, and any risks associated with the potential identification of any one patron through the publication of loan data. The second area of concern relates to the functionality and feasibility of the idea itself, involving how and where the physical repository of the collective open datasets would be hosted, the ongoing financial implications of hosting the data, and the ongoing security and preservation of the data overtime. Each of these areas of concern will be explored in turn.

Privacy principles and legislation

Privacy and the Australian law

Publication of any personal user data necessitates a thorough assessment of the risk such publication poses to the privacy of any one individual or group of people. A survey of the legal frameworks protecting the privacy and data of individuals within Australia is a critical step towards adequately addressing this key concern in relation to the feasibility of the publication of open data across the Victorian library sector.

Within Victoria, public sector organisations are bound by the Privacy and Data Protection Act 2014. This legislation determines that interference with the privacy of an individual can be identified if the “act or practice [in question] is contrary to, or inconsistent with [either] an Information Privacy Principle or an applicable code of practice, a public interest determination...an approved information usage arrangement, or a current certificate.” (Privacy and Data Protection Act, 2014, p. 39). Simply put, publication of an open data set of Victorian library borrowing and collection statistics must not interfere with any of the 10 Information Privacy Principles set out in the usage agreement with the collection management system that hosts the data or in the relevant legislation. The usage agreements for all collection management systems involved within this study determine that the library service (user) has proprietary ownership over the data collected. As such, use of this data does not constitute a breach. However, further investigation of potential breaches of the Information Privacy Principles is required.

The 10 Information Privacy Principles contained within the Privacy and Data Protection Act oblige organisations, including libraries, to handle all personal information in accordance with a set of principles that pertain to the following areas:

1. Collection
2. Use and disclosure
3. Data quality
4. Data security
5. Openness
6. Access and correction

7. Unique identifiers
8. Anonymity
9. Transborder data flows, and
10. Sensitive information¹

Principles 2 and 4 are of particular pertinence when considering the risks associated with the publication of open data across Victorian libraries. Principle 2, related to use and disclosure, states that “an organisation must not use or disclose personal information about an individual for a purpose (the secondary purpose) other than the primary purpose of collection” unless and until a particular set of extenuating circumstances apply (Office of the Victorian Information Commissioner, 2023b).

In order to determine whether the publication of the Victorian library sector's data would be in breach of this principle, it is critical to first understand what is meant by “personal data”. The Office for the Victorian Commissioner determines information to be “personal” in nature if “the individual's identity [is] apparent or reasonably ascertainable”, clarifying with the following:

An individual's identity will be apparent if it is clear from the information itself. For example, if their name and other identifying information is directly connected to the information. An individual's identity is reasonably ascertainable if it can be determined by taking reasonable steps. For example, could other information be combined with the information in question that would lead to the individual's identity becoming apparent? (Office of the Victorian Information Commissioner, 2023a)

Thus, a rigorous data de-identification process must be implemented in order to be assured that a user's identity was not immediately or reasonably apparent.

De-identification explained

The term “de-identified” describes data that cannot be used to either immediately or reasonably identify a person. This data can be deidentified from the time of collection, or be manipulated to exclude certain fields that would make it identifiable (Office of the Information Commissioner, 2023b). Once data can be reasonably considered deidentified, it is no longer subject to the principles contained within the Privacy Act 2014 (Office of the Information Commissioner, 2023b).

To ensure this, the whole of the data situation must be assessed. In other words, both the data itself and the environment in which it will be received need to be considered, both individually and together, in order to determine whether true deidentification has been achieved or is indeed possible (Office of the Information Commissioner, 2023). For example, a dataset that includes the gender, age range, and postcode of an individual would not be reasonably considered re-identifiable within a large population sample such as postcode 3000, which encompasses Melbourne CBD. However, this data may be considered reasonably re-identifiable within a smaller population, such as a small area with few residents. In this way,

¹ Further information about each of these principles can be accessed online via the Office of the Information Commissioner at <https://ovic.vic.gov.au/>.

deidentification of data can be seen as “a risk management exercise, not an exact science” (Office of the Information Commissioner, 2023b).

Risk assessment

A risk assessment for the open publication of Victorian library sector data must consider two factors: how the data can best be modified, if necessary, to avoid reidentification, and whether this data is still of utility once this has been achieved. There is little use in publishing data that is no longer robust enough to support the use cases detailed above. In undertaking this feasibility study, multiple methodologies for safely and appropriately deidentifying personal data, whilst retaining functionality, were considered and in practice did not create any significant challenges to collecting useful data. The LMS reporting software used in the feasibility study allowed the creation of reports to provide the target data without exposing any personally identifiable information.

Should this not be the case for any particular dataset, CSIRO, in partnership with the Office for the Information Commissioner, have produced a deidentification decision making framework to assist in this process. This framework provides a set of techniques that allow data custodians to identify, assess, quantify and mitigate the benefits and risks involved in the release of open datasets (CSIRO, 2019).

Risk management is a contextual and ongoing process, requiring consistent assessment and reassessment of both the data and the data context in order to enable the release of information that is both safe and useful. For this reason, it is not within the scope of this feasibility study to undertake a full risk assessment. However, the project team recognise the significance of this part of the process, reflected in the recommendations at the conclusion of the study.

Data hosting and security

For an open data initiative of the type being investigated in this paper to proceed some practical considerations need to be addressed including how data is hosted, published and managed.

Server location

An open data initiative requires that the dataset be hosted on a server connected to the internet, allowing unrestricted access to the data by the public. Servers, while traditionally a machine physically located on premises, controlled and supported by an organisation's IT teams, are increasingly moving to cloud based services.

On premises servers provide an organisation with greater control and security over access to the data, but requires a larger upfront cost for the purchase of hardware, software and associated infrastructure associated with the server, as well as ongoing staffing resources to support the servers. Scalability of on premises servers can be limited both by budget and physical limitations of on premises data centres.

Cloud based services provided by a third party vendor are located away from the organisation's premises. Responsibility for server security, maintenance and reliability is placed in

the hands of a vendor. Organisations need to be satisfied that the vendor will comply with necessary standards for security and support prior to engaging their services. Public sector organisations need to also give consideration to the physical location of the data warehouses, ensuring that it is housed within Australia and ideally in Victoria. Cloud based services have the advantages of lower up-front costs and a payment model that allows users to pay for current usage rather than anticipated need. The services are scalable, able to grow with the size of the dataset and usage, and responsibility for updates, compliance and security are delegated to the vendor, and do not translate into a resourcing requirement of the organisation.

While existing on premise server infrastructure may exist among participating library services, based on the potential growth of the dataset year on year, and potential to expand to include historic data, the scalability, support and pay as you go approach provided by a cloud based service is attractive. Before commencing a detailed analysis should be entered into to fully understand cost implications.

Data publishing and platforming

While this project is primarily giving consideration to making collection usage data open and accessible to all, further thought should be given to where data is published in order to make it most accessible and visible to potential users. It would be anticipated that this would include existing library services online presence, but could also be made available through platforms with greater reach including services such as the Victorian Government run DataVic platform.

Furthermore, if an open data initiative like this were to progress, consideration should be given to platforming that supports users to manipulate and gain insights from the data in the platform while also allowing downloads of the data file for manipulation in preferred software.

Data management

The feasibility of an open data initiative aggregating collection usage data from Victorian public libraries is dependent on support and cooperation of Library Managers, IT and collections teams from participating library services in the planning, execution and maintenance of the data sharing process.

While this paper doesn't make specific recommendation around the method for sharing data by each library services, it is noted that there are multiple potential options based on the frequency with which the data is collected, the capability/compatibility of the diverse software in use, the frequency of data uploads, and potential for automation of data collection using APIs. All options have potential staffing and cost implications and will rely on a level of development and data management to ensure the quality and accuracy of the output. It is noted that across the public libraries and State Library Victoria there are a number of data engineering roles performing parallel data management activities. The introduction of management for a new dataset to existing roles could be feasible with limited impact on ongoing workload.

Similarly, library services having existing reporting processes/obligations aligned to the requirement of this proposal should mean that the addition of this process would not prove burdensome for participants in the long term.

Evaluation

Stakeholder consultation

Methodology

Stakeholder engagement was undertaken firstly by identifying key groups who would have a potential vested interest in the creation and use of the open datasets. and then speaking to key individuals within those identified groups as described below:

- Statewide library sector
- Victorian public library staff
 - Library service managers (metro, regional and corporation)
 - Collection managers
 - IT managers
- Suppliers and vendors (including of e-resources)

Individuals were interviewed to determine their understanding of open data principles, address concerns or misgivings they may have, and most importantly ascertain what benefits or application they foresee with the development and publishing of the open dataset. The questionnaire and the raw data collected during stakeholder interviews can be referred to in Appendix 1, attached.

Stakeholder theme 1: how can I protect myself from financial and reputational risk?

It was identified during stakeholder engagement with eresource platform providers (e-book and audio book) that it is of utmost importance that the provider is not identifiable in order to protect their own business model from competing eresource vendors. This would hold true for all suppliers even for physical items (i.e. it is the cumulative circulation of the item that is being reported on, regardless of the supplier or source of supply). Of course this then extends to the reputation of individual library services, and moreover to the reputation of publishing houses and authors.

In consideration of the financial and reputational risks the following questions arose:

- What does it mean to be identified as the lowest circulating author across all Victorian public libraries?
- What does it mean for individual Victorian public libraries if their circulation statistics are lower than other libraries comparable in size and with a similar demographic (for example would it affect future funding, budget bids, etc ...)
- What does it mean if the lowest circulating items identified were consistently from the same publishing houses?

When considering open data, it is important to note that concerns such as the above will always be raised, many of which cannot be adequately addressed within the scope of this

project. It is not the intention of this feasibility study to dispel every concern of all stakeholders, but rather to highlight the benefits such a project could bring to the library sector and beyond.

Stakeholder theme 2: how will the data be used and by whom?

Examples of the possible application of the open datasets within the library sector have been identified above in the section regarding **Use Cases**. The benefits of open datasets in relation to decision-making regarding collection development, promotion and management, assisting with strategic direction and future investment including improved and transparent reporting outcomes has been established.

An additional common theme identified via the stakeholder engagement posits *what is the benefit of having the datasets publicly available?* A common question that arose was “how would the datasets be used and by whom outside of the library sector?”

Many of the stakeholders interviewed raised the potential misuse of published data by members of the public or lobby groups as a concern. One stakeholder observed that the data could be used by members of the public running for council or for other political ends. Although it is prudent to acknowledge this concern, it is important to note that the risk of misuse of library data does not exponentially increase with the creation and public access to open datasets. Library data already exists in publicly available formats such as library catalogues (library holdings), Public Library Victoria Annual Reports, and Council or Library Service Annual Reports. It was also noted that transparency of data and reporting can conversely increase perceptions of confidence and integrity: if you are looking at a raw data source, or know it is readily available, you can have greater confidence that reporting has not been manipulated, or data ‘cherry-picked’, to reflect a particular point of view. As one stakeholder also acknowledged, the risks are equally weighed against what the potential benefits that may arise from others having access to the data may bring, ‘If we are not asking good questions about our data, perhaps others will’.

In the **Use Cases** above, two additional groups have already been identified as potentially benefiting from the existence of the open datasets being a) authors, publishers and suppliers and b) state and federal government agencies.

The reality is that the extent of the application or use of the data and by whom is unknown. In many ways this is why the creation of this open data set is such an exciting and important initiative.

Stakeholder theme 3: how will this be resourced, both in staffing and funding

Conversation around this new initiative inevitably turned to the cost of such an undertaking and the resourcing required for its success. While stakeholders overwhelmingly supported making collection usage data open and accessible, it was unsurprising, particularly in the current economic context, that the hard nosed reality of who would do the work, where the funds would come from to support the initiative was raised. While it isn't in scope of the project to make recommendations around the mechanism for funding this initiative or by whom, discussions with stakeholders highlighted a few key take-aways.

- Data storage is cheap - The publication of this open data set is at its most basic the publication of a file containing several million lines of information that will grow year on year. Based on 21/22 figures this would be in the order of 43.5 Million lines totalling 4.4 Gb in size, that could grow by this figure each year. Whilst seemingly a large number, in the world of data this is seen as a modest figure and not a significant financial burden to store and maintain.
 - Cloud based hosting services such as GitHub, currently used to host select State Library Victoria datasets, and AWS offering free service for eligible public repositories and open datasets (GitHub, 2023; AWS, 2023). Paid services price storage of data at around \$0.025 per GB per month, meaning storage of this data set annually would be in the order of \$6 (AWS, 2023).
 - While there will no doubt be other costs associated, it highlighted that data storage costs are unlikely to prove a limiting factor for the project.
- Automation of process should be the goal - Participating library services will need to commit staff resources to ensure the success of the project. Were the initiative to receive broad support from the sector, initial work would need to focus on the mechanisms through which data is extracted and shared from LMSs of participating library services. In discussions with stakeholders the potential to automate processes through APIs, while requiring an upfront investment of staff resources from each service, was highlighted as a mechanism that should be explored to reduce ongoing demands on staff time and open possibilities to increased frequency of data updates. It was also noted by stakeholders that within PLV library services and State Library Victoria there are staff performing aligned activity for who are equipped and energised to support the project.
- Ownership for the data set should be established - stakeholders identified that the chance for success of the initiative would be enhanced if responsibility for the data set was clearly established and agreed, with responsibility for processing, any manipulation of data and the ongoing maintenance of the open data set tied to a specific role. Experience from previous open data initiatives, while strongly supported in early phases, have suffered longer term from being the passion project of one staff member, and as a result have been neglected when that person leaves the institution. Tying responsibility to a role should help to ensure the longevity, accuracy and therefore value of the data set.

As highlighted in previous sections, across the public libraries and State Library Victoria are a number of data engineering roles performing parallel data management activities to which there are natural synergies with this activity. While not making recommendations on where responsibility for this activity should reside, we noted from discussions with stakeholders in these roles the enthusiasm and openness to supporting such an initiative and the lack of concern for the workload this activity presents.

Stakeholder theme 4: advocacy and stakeholder engagement is critical.

Despite the concerns raised by the stakeholders engaged throughout the feasibility study, the data overwhelmingly indicates a net positive attitude towards the project, with many interviewees expressing interest in the potential applications of open data within the library sector. Further advocacy within individual library services, as well as across the sector as a whole, will be critical to the implementation and success of the proposal. Key figures within the sector should be engaged, including library CEOs, PLV executives, and ICT staff across the state.

Sample data set and application

Library services surveyed

Hume Libraries	SirsiDynix Symphony - Libraries Victoria Consortium
Wyndham Libraries	SirsiDynix Symphony - Libraries Victoria Consortium
City of Greater Dandenong Libraries	SirsiDynix Symphony - Libraries Victoria Consortium
La Trobe City Libraries	SirsiDynix Symphony - Libraries Victoria Consortium
Indigo Libraries	SirsiDynix Symphony - Libraries Victoria Consortium
YPRL	Civica Spydus - Data supplied to project group

Data collected - loans report

Data Field	Description	Notes
Catalog Author	Author as described in bibliographic record	No issues - easily matched
Catalog Title	Title as described in bibliographic record	No issues - easily matched
Item Type	Book, DVD, Magazine, etc	No issues - easily matched
Item Collection	Fiction, Non-Fiction, etc	YPRL combine Item Collection and Item Audience into a single field. In our sample this information was split manually in bulk.

Item Audience	Target age group for the item - Adult, Junior, Teen	YPRL combines Item Collection and Item Audience into a single field. In our sample this information was split manually in bulk.
Item language	Language of the item as described in the Bibliographic record	
Item Library Service	Owning library of the item	It was necessary to create a new column in the data supplied from YPRL for this field.
Item Library Service type	Metropolitan or Regional	It was necessary to create a new column in the data supplied from YPRL for this field.
Item Created Month	Month the item was created in the LMS in numeric format 1-12	
Item Created Year	Year the item was created in the LMS in 4 digit format	YPRL provided the exact date so the year in a 4 digit format was derived manually
Loan library	Library branch where the loan took place	
loan or renewal	Does the transaction row represent a loan or a transaction	
Loan Date	xxxx-xx-xx	
Loan Month	Month the item was loaned in the LMS in numeric format 1-12	In our sample this is always '5'
Loan Year	Year the item was loaned in the LMS in 4 digit format	In our sample this is always '2023'
User Age	numerical age 0-120	Data from YPRL contained only user birthday, requiring manual bulk reformatting before matching.
User postcode	4 digit format	Our sample dataset does not feature matched postcode as a result of a formatting error in the matching attempt.

Data collected - holdings report

Data field	Description	Notes
Catalog Author	Author as described in bibliographic record	No issues - easily matched
Catalog Title	Title as described in bibliographic record	No issues - easily matched
Item Call Number	Call number as described in bibliographic record	No issues - easily matched
Item Collection	Fiction, Non-Fiction, etc	Generally uniformly described
Item Audience	Target age group for the item - Adult, Junior, Teen	YPRL combines Item Collection and Item Audience into a single field. In our sample this information was split manually in bulk.
Item Type	Book, DVD, Magazine, etc	
Item language	Language of the item as described in the Bibliographic record	Generally uniformly described
Item Created Year	Year the item was created in the LMS in 4 digit format	YPRL provided the exact date so the year in a 4 digit format was derived manually
Item Library Service	Owning library of the item	
Item Library Service type	Metropolitan or Regional	
Item Price	AS described in the bibliographic record	
Item Legacy Total Charges	Total lifetime loans on the item	

Data collected - size

Sample Dataset	Rows	Size
Example Loans dataset	422,872	40mb
Example Holdings Dataset	1,059,442	62mb

Based on reported statistical measures for 2021-22, we can estimate the size of the annual data set, with the loans and holdings information from Victoria's 50 public library services.

Projected Annual Dataset	Rows	Size
12 months of loans, 50 library services	34.8M	3.48gb
Collection holdings 50 library services	8.67M	867mb

Discussion

The majority of the data in this sample set was derived from reports within BlueCloud Analytics, the reporting software for the Libraries Victoria consortium. In order to provide insight into the merging of data from different systems, data for both datasets was provided by Yarra Plenty Regional Library service, exported from Civica's Spydus LMS product. These six library services represent a rough cross section of the types of library services that exist in Victoria - 2 small regional libraries (La Trobe, Indigo), 3 larger metropolitan single council library services (Hume, Wyndham, CGD), and 1 regional Library corporation (YPRL). We did attempt to obtain data from a second non-LV library service from a third LMS but ultimately didn't obtain such a dataset.

Combining the loans and holdings data from the 5 Libraries Victoria consortium members was relatively straight forward, as each library service uses the same fields to describe items and user information, the target data could be collected by running a single report for each of the two datasets. There are some distinctions between the way each field is actually used within each library service, but not in a way that significantly impacts the value of the data.

Challenges

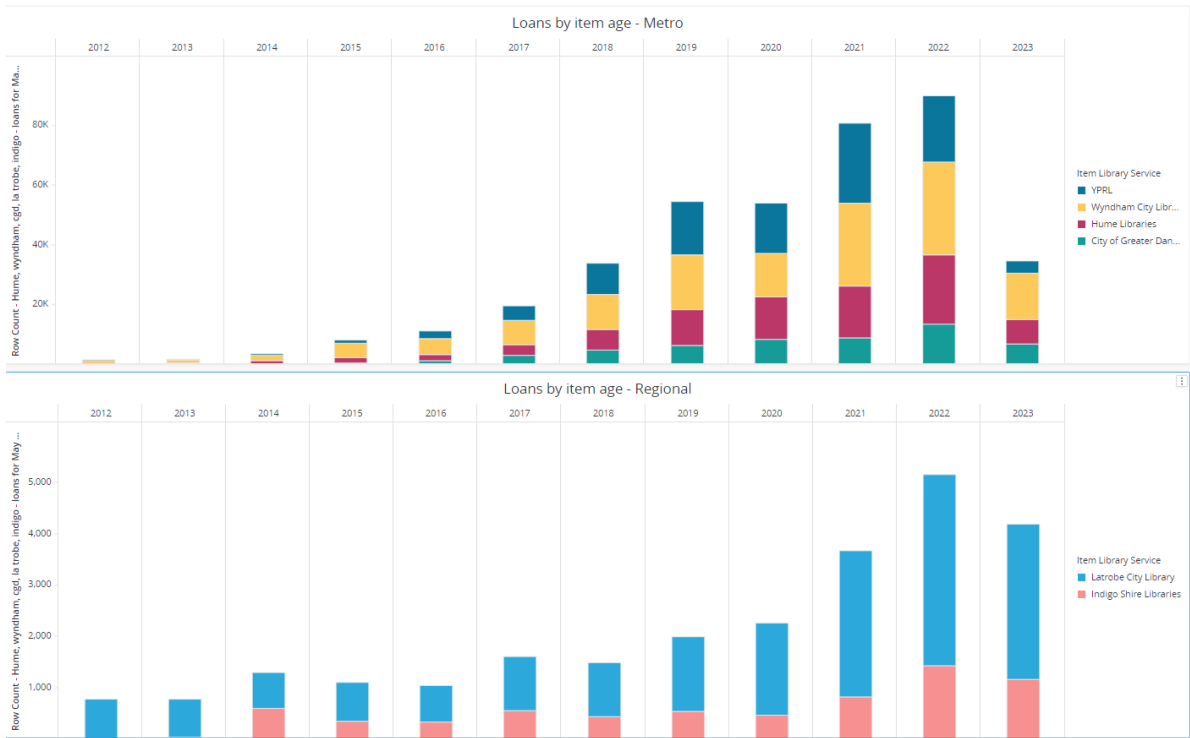
Combining the consortium data with the data provided by YPRL did create some challenges as although the same information was provided there were formatting differences that required some data manipulation before it could be usefully matched with the main data set. This points to one of the biggest challenges for a large-scale implementation of this project, creating a useful dataset with the data from multiple LMSs would require each library service to provide their data in exactly the right format to be easily matched.

There were other observations from the data matching exercise that point to further challenges for creating a clean dataset, such as irregularities in loan figures recorded by central processing offices compared to their branch libraries. Ideally the sample loans dataset would have

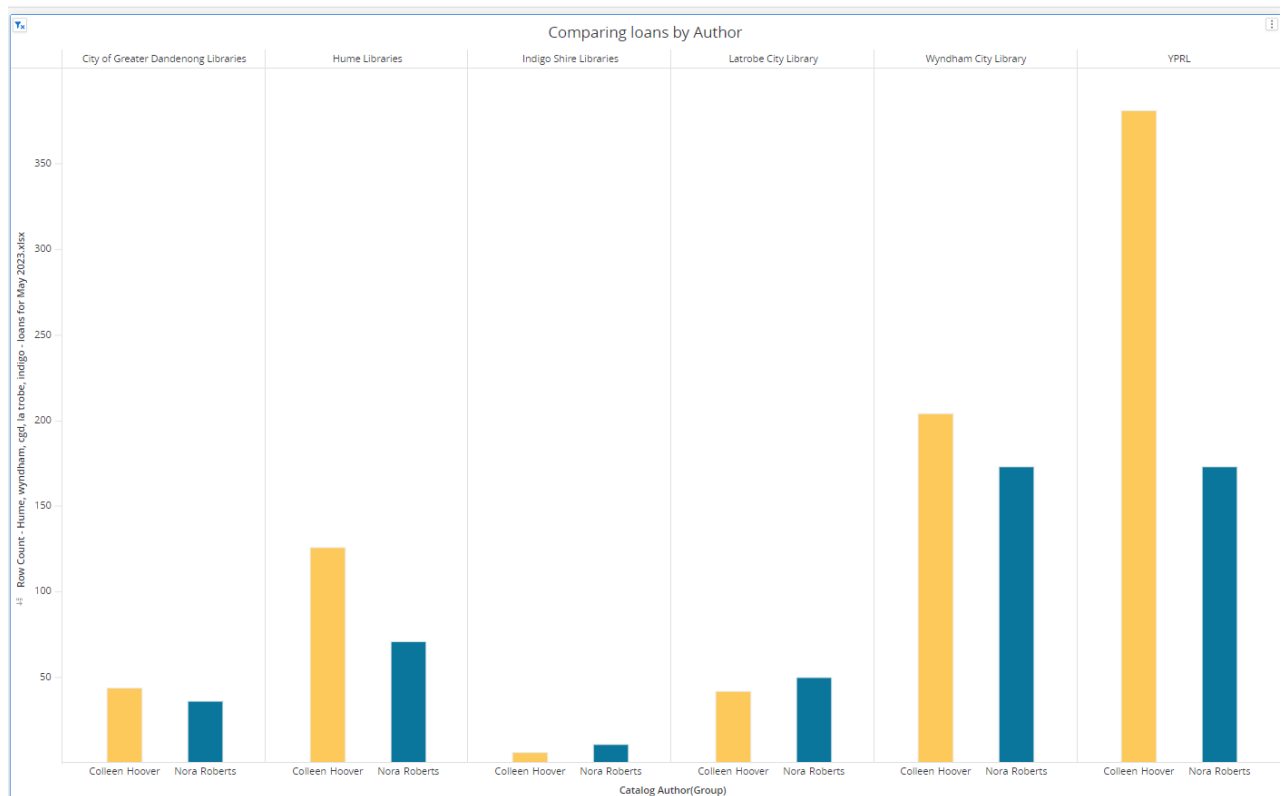
contained 12 months of data so analysis could be conducted over that time period however the size of the reports that these created meant that it wasn't practical given the available tools and skill set available to the group.

Data analysis examples

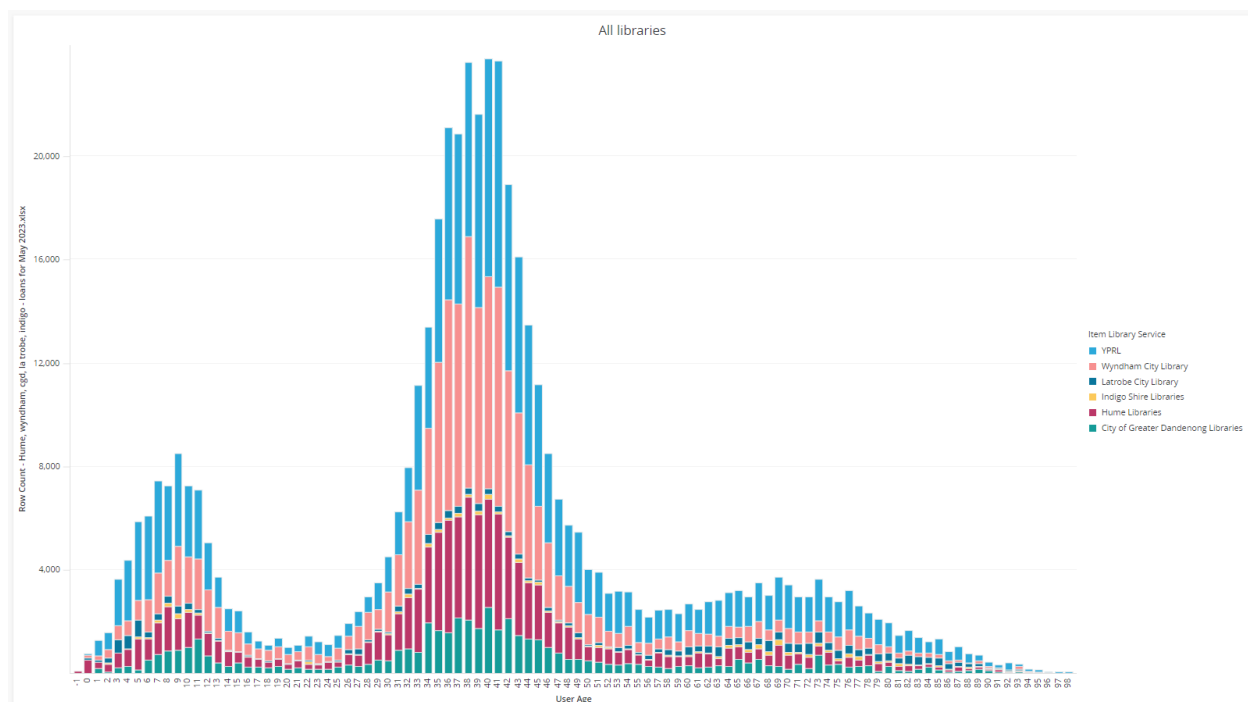
In the below chart we can see clearly how metropolitan libraries are able to keep their collections developed with newly published items, whereas regional libraries continue to rely on older materials for a higher percentage of their loans.



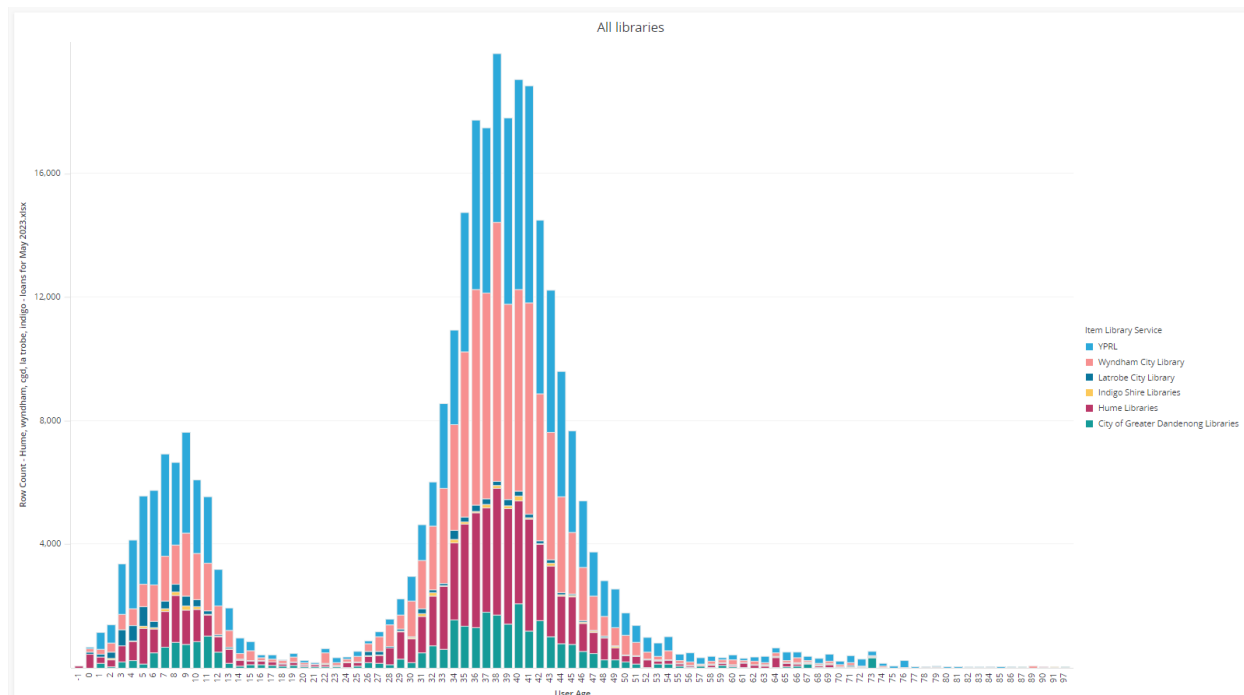
This chart is an example of the very granular level of comparison we are able to conduct with these datasets, looking at the loan performance of two popular authors. Nora Roberts, a popular mainstay of the romance genre and Colleen Hoover, who has exploded in popularity thanks to social network TikTok.



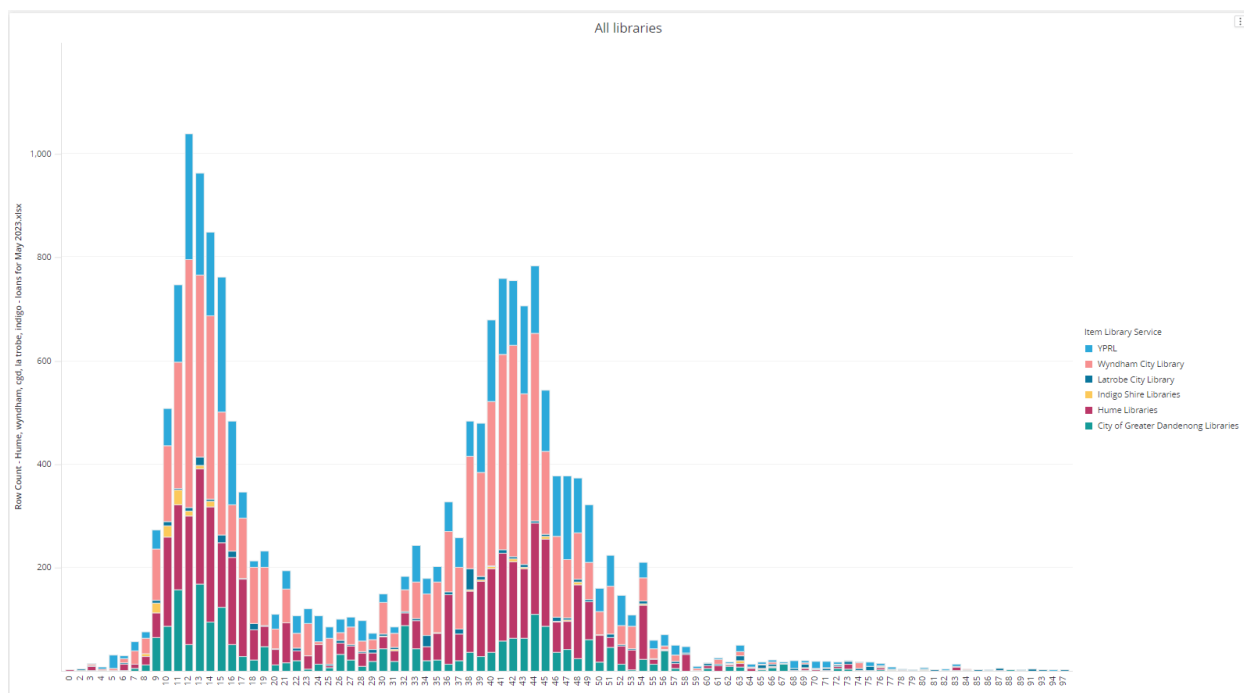
Looking at the user age field we can get a picture of the age profiles of public library users.



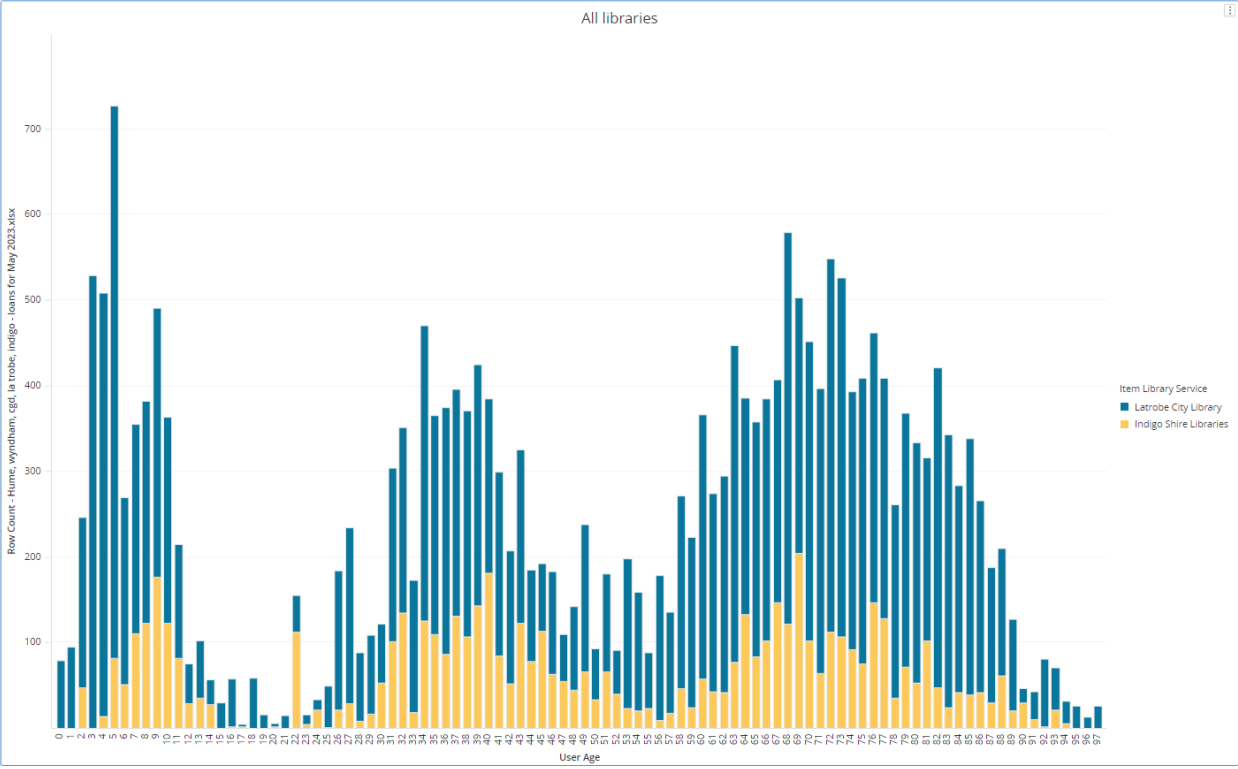
When this chart is filtered to only include junior items, we can see that a large percentage of loans are happening on adult membership cards. While not surprising, this does suggest that junior members could be under-represented in reported membership statistics.



Filtering the chart to include only Teen items, we get a different story, which perhaps confirms what we already know, that there is a large audience for YA fiction among the 30-40 age group.



Looking just at regional libraries, we see a very different spread of ages using the library, with much more in the 60+ age group and very small numbers in the 15-21 age group.



Recommendations

It is recommended that:

- The collation and publication of an open data set of Victorian library service loan and collections data is both feasible and necessary for the growth and development of the sector. A move towards open data as standard will demonstrate the sector's commitment to transparency and democratic access to information.
- Further education and advocacy work is undertaken across the Victorian public library sector, allowing staff to further familiarise themselves with the concept and the potential benefits. Recommended methods of advocacy include:
 - The distribution of this feasibility study amongst members of the ICT SIG, accompanied by a presentation from the project team on the findings
 - Agreement with PLV to create a pilot system, with a dashboard interface, that can be accessed by public library staff to grow comfort with the concept and foster confidence in evidence-based decision making
 - Recruitment of interested library services to volunteer their own data, i.e. through a .csv file to enable this pilot

References

Amazon Web Services.

<https://aws.amazon.com/pricing/>

Australian Curriculum and Reporting Authority. (2023). My School.

<https://www.myschool.edu.au/>

Australian Government. (2015). Australian Government Public Data Policy Statement.

https://www.finance.gov.au/sites/default/files/2022-10/aust_govt_public_data_policy_statement.pdf

Biswas-Diener, R., & Jhangiani, R. (Eds.). (2017). *Open: The philosophy and practices that are revolutionizing education and science*. Ubiquity Press.

Chignard, S. (2013). A brief history of open data. *Paris Tech Review*.

<http://www.paristechreview.com/2013/03/29/brief-history-open-data/>

City of Greater Geelong. (2023). Geelong Data Exchange – making data user friendly.

<https://www.geelongaustralia.com.au/smartcity/article/item/8cfbdf8c330f3e1.aspx>

City of Melbourne. (2020). Urban Forest Visual. <http://melbourneurbanforestvisual.com.au/>

City of Melbourne. (2023). Open Data.

<https://www.melbourne.vic.gov.au/about-council/governance-transparency/open-data/Pages/open-data.aspx>

City of Yarra. (2018). *ICT – Open Data Policy*.

<https://www.yarracity.vic.gov.au/-/media/files/events/council-and-pdc-meetings/2018-meetings/council-meeting-22-may-2018/1121--open-data-policy--attachment-1.pdf>

CSIRO. (2019). Engineering identity from anonymity: Our work on risks of re-identification.

<https://algorithm.data61.csiro.au/engineering-identity-from-anonymity-our-work-on-risks-of-re-identification-3/>

GitHub.

<https://github.com/pricing>.

International Federation of Library Associations and Institutions. (2021, December 15). IFLA Statement on Open Library Data [Press release].

<https://repository.ifla.org/bitstream/123456789/1842/2/IFLA%20Statement%20on%20Open%20Library%20Data.pdf>

Office of the Victorian Information Commissioner. (2023a). Key Concepts.

https://ovic.vic.gov.au/book/key-concepts/#Personal_information

Office of the Victorian Information Commissioner. (2023b). Privacy law - an overview.

<https://ovic.vic.gov.au/privacy/resources-for-organisations/privacy-officer-toolkit/privacy-law-an-overview>

Open Knowledge. (n.d). *What is open data?*

<https://opendatahandbook.org/guide/en/what-is-open-data/>

Privacy and Data Protection Act 2014 (Vic).

<https://www.legislation.vic.gov.au/in-force/acts/privacy-and-data-protection-act-2014/027>

Productivity Commission. (2017). *Data availability and use* (Report no. 82), Canberra.

<https://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf>

State Government of Victoria. (2020). *Data policies and standards*.

<https://www.vic.gov.au/data-policies-and-standards#address-data-standards>

State Government of Victoria. (2021a). *DataVic Access Policy Guidelines: Introduction*.

<https://www.data.vic.gov.au/datavic-access-policy#datavic-access-policy>

- State Government of Victoria. (2021b). *Policies and standards for government IT*.
<https://www.vic.gov.au/policies-standards-for-government-IT>
- State Government of Victoria. (2021c). *DataVic Access Policy Dataset Publishing Manual*.
<https://www.vic.gov.au/datavic-access-policy-dataset-publishing-manual>
- State Government of Victoria. (2023). *Crime Statistics Agency*.
<https://www.crimestatistics.vic.gov.au/>
- State Government of Victoria. (n.d.). *Homes Victoria [Datasets]*.
<https://discover.data.vic.gov.au/organization/homes-victoria>
- State Library of Victoria and Public Libraries Victoria Network. (2013). *Victorian Public Libraries 2030 Strategic Framework*.
<https://www.slv.vic.gov.au/sites/default/files/VPL-2030-strategic-framework.pdf>

Appendix 1: Stakeholder engagement mapping

Tier 1	Tier 2	Tier 3	Date	Engagement
Jackie Horwood Acting Head of LSE, SLV			31/3/2023	Kate McHugh Karmel Savahl
	Leanne Williams CEO, MyLi		10/03/2023	Karmel Savahl
	Susie Prestney Manager Libraries & Arts Melton City Council		25/03/2023	Karmel Savahl
	Jenny Mustey Library Services Manager – Campaspe Regional Library		25/03/2023	
	Monique Godbehere Library Services Senior Coordinator – Greater Dandenong Library		23/05/2023	
		Gareth Evens Content Management Librarian - Greater Dandenong Libraries	26/5/2023	
Julianne Tobin Account Manager Overdrive (Libby)			10/03/2023	Karmel Savahl
	Monika Szunejko Director - Collections State Library Victoria		31/05/2023	Kate McHugh Chris Nye
Paul Duldig CEO State Library Victoria			15/06/2023	Kate McHugh Chris Nye Patrick Jovaras
		Mathew Plachotny Library Systems & Technology, Coordinator, Wyndham City Libraries		Jenny McKnight
		Aidan Murphy Collections & Information Services Coordinator Wyndham City Libraries		Jenny McKnight

Questionnaire

1. What is your understanding of open data and open data principles?
2. Do you feel there is a benefit to enabling public access to library collection usage statistics?
3. If yes, why?
4. If no, why not?
5. What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?
6. In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?

Stakeholder engagement interviews

Stakeholder 1	
<i>Who are the other relevant stakeholders who should be approached regarding the project?</i>	Start with PLV executive team. Suggested getting a cross section of Library managers e.g. Corporation, Single Municipality, Regional Library. Perhaps start with: Leanne Williams (Mylly), Susie Prestney (Melton), Jenny Mustey (Campaspe)
<i>What were your thoughts when you first heard of the project? Do you feel there is a benefit to enabling public access to library collection usage statistics?</i>	Yes – believes libraries should be transparent and share data.
<i>What were your concerns or misgivings about the project? What more information do you think is required to make the sector more comfortable with the concept of library collection usage statistics open datasets?</i>	Data would need to be de-identified. Concerned regarding privacy and security. Also concerned about the misuse of data. Gave example of misuse of medical data during covid.

Stakeholder 2	
<i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i>	Consider "why?" What is the cost benefit of having this information available in this format? What if the info goes out and not one ever uses it? Is this concept innovative and is it trying to solve a specific problem? Would this assist staff in making decisions – or do they have the information already to hand that they need.
<i>If no, why not?</i>	<p>Concerns about deidentifying data – even if you deidentify it – then it means that you are collecting data in the first instance that didn't necessarily need to be collected. Also ethical concerns about how the data could be misused.</p> <p>What is the return on investment? Is the actual concept innovative – does it provide business benefit, or make efficiencies?</p>
<i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i>	It isn't about the open data concept. Questioning the value contribution to the beneficiary.
<i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i>	Acknowledges that different members of the organisation would have a different take. Different library services would have a different take. Speak to other library services and potentially publishers.

Stakeholder 3	
<i>Do you feel there is a benefit to enabling public access to e-resources library collection usage statistics?</i>	In theory yes, as exists for physical but no equivalent for digital. Overdrive is always keen to provide information if it is for the "greater good".

<p><i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i></p>	<p>Initial concerns about competitors getting information, however if combined line items then not distinguishable.</p> <p>Happy for authors to get remunerated.</p> <p>Would be good for patrons and other library services to see what is trending and what needs to be held for collections to be up to date.</p>
---	--

Stakeholder 4	
<p><i>What is your understanding of open data and open data principles?</i></p>	<p>My understanding is based on the principle that the data should be shared and it is aligned with our principles of open knowledge sharing and the free-flowing of information.</p> <p>From a business perspective it provides you with a strong foundation for your business intelligence as it allows for comparative data and benchmarking.</p> <p>A further benefit is that it allows for transparency and integrity of data. If you are looking at a raw data source you know that it has not been manipulated to reflect a particular point of view.</p>
<p><i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i></p>	<p>Yes.</p>
<p><i>If Yes, why?</i></p> <p><i>If no, why not?</i></p>	<p>I believe so. We have an organisational commitment to transparency and the public can benefit from this transparency in terms of their trust in public institutions and accountability around allocation of how resources are allocated and spent. To illustrate with an example, an imaginary under-resourced library service should be able to see that their collection is smaller or older than other services.</p>

<p><i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i></p>	<p>I think probably what I'd need to know is how we would handle any concerns raised around that data. Open data can raise difficult questions for organisations and those organisations may not be well-placed to respond to those questions.</p> <p>I would need to understand the processes behind each source before it is made openly accessible. It could be that the data is manipulated to make it look better or it could be just the case that data is recorded in different ways.</p> <p>The precedent is already there in that we already share the stats on the PLV website—if anyone was going to get up in arms about it they can do this. This dataset improves things in that it makes things far more granular and potentially live.</p>
<p><i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i></p>	<p>PLV would be a key stakeholder in that they are the owner of the library statistical reporting.</p>
<p><i>Now you've seen some of our sample datasets, is there anything else you'd like to add?</i></p>	<p>I love it. There are things I can see here that I would use in my day-to-day job.</p>

Stakeholder 5	
<i>What is your understanding of open data and open data principles?</i>	It's different across government authorities, but the Australian Government has an expectation that data be made open if possible and that it's made open under specific licences.
<i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i>	I'd say 'yes', but I don't know what that value is. It's valuable amongst librarians and library staff, but then again, I don't know how it would be used elsewhere. That's the whole wonderful thing about open data - you don't know how things are going to be used. If we are not asking good questions about our data, perhaps others will.
<i>If Yes, why?</i> <i>If no, why not?</i>	It's always a possibility that members of the community could use it for bad ends. Even potential councillors could go through the data and use it for political purposes.
<i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i>	<p>I would really want to know exactly what data will be made open. A clear example needs to be provided.</p> <p>The only thing in the sample datasets that made me a bit anxious was the Stella Award analysis as that implied there might be line-by-line analysis; also having this level of granularity would also result in a huge amount of data to be stored and managed. This is selfish, I have to ask (as IT Coordinator): how much work is there going to be for me? What will be the risks? Will there need to be a privacy impact assessment?</p>
<i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i>	<p>Michelle McClean (for the PLV lens)</p> <p>Privacy Officers across various organisations</p>

Stakeholder 6	
<i>What is your understanding of open data and open data principles?</i>	Not familiar with the breadth and width of what the data can show. Once you have access to it, what are you able to do with it? Are we using our own data? Who will see our data?
<i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i>	Yes, but interested to know who would be using it?
<i>If Yes, why?</i>	Can see how it is beneficial to libraries, library students, maybe publishers and authors. A concern though - could something be picked up and be used by the media for good or bad? Can see the benefits of it but feels she needs an understanding in a broader context. She sees it would be a good benchmark comparative tool as long as the data is correct. We talked about the LOTE aspect and it being beneficial in starting a new collection - could use as a tool to gather titles and authors for a particular LOTE language, targeted to a like demographic.
<i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i>	<p>Comfortable with the concept - however still has questions regarding application, especially outside of a library service setting. could see that potentially could be used for thesis or academia.</p> <p>What does practical usage and intent look like (after seeing the examples). What would be the cost to the service and would the cost sit with the service (in which case service would need to be actively using the tool).(ROI)</p>
<i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i>	Governing bodies, e.g. ALIA, PLV.

Stakeholder 7	
<i>What is your understanding of open data and open data principles?</i>	Basic understanding of shared data and shared usability.
<i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i>	Yes for all library staff. Can see the applicability for Collection Development and how it can assist for profile buying.
<i>If Yes, why?</i>	Shared data is important for advocacy - being able to see in a geographic sense. Can change the way libraries display/promote their items (referring to the example that over 35 yo borrowing Young Adult material). Even how to market resources through social media.
<i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i>	Is comfortable with the concept. No issues with libraries using as long as there is no lens that compares performance or bias written into it. However if it is raw statistics then it is objective. Can see how library services could potentially use the data if they don't have the skillset to analyse their own data - it provides evidenced based. Also sees the value in LOTE - creating collections and getting an understanding.
<i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i>	<p>Vendors would find this interesting to know what to supply.</p> <p>Likes it as it is evidenced based and can allow libraries to meet community needs and be ahead of the game.</p>

Stakeholder 8	
<i>What is your understanding of open data and open data principles?</i>	Anyone can use datasets to create a picture and tell a story in a visual form. Mashup up information and make it usable for decision making.

<i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i>	Yes.
<i>If Yes, why?</i>	However - not sure how popular it would be outside of library use. Can see it can be used for advocacy (positive and potentially negative if the data is misused). Can see potential use for Publishers and planners. Currently is creating library strategy and needs to engage a consultant - this information would be beneficial to assist.
<i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i>	Overall if people understand that the data tells a story - it can be localised to your service or used for comparative/benchmarking. It's evidence based.
<i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i>	General library staff.

Stakeholder 9	
<i>What is your understanding of open data and open data principles?</i>	Datasets that are in the public domain that can be accessed for whatever usage the user would like.
<i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i>	Yes, it provides an opportunity for bigger picture analysis of trends but also gives an opportunity for people outside of the sector or new to the sector to get access to the data.
<i>If Yes, why?</i>	Transparency. See relevance for Acquisitions - if you can demonstrate Collection Acquisition that is commensurate to demand. Accountability - could stop the politics of lobbying for items / collections that don't move (conversely gives credence to the introduction of new collections and funding)

<p><i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i></p>	<p>Questions:</p> <p>Risk averse and concerned about deidentified data (privacy).</p> <p>Cost Benefit analysis - What is the ROI. Who is hosting and coordinating? Who is managing/sweeping data so that it is uniform and comparable?</p> <p>What happens if not all libraries engage or drop out over time (usability of data)?</p> <p>Is there potential for the data to be misleading because there is no context (just data, no story) or is this not a problem or an end user issue only? Visualisation of the graphs peaked interest, but would this be available in the dataset or is it just line item data available?</p>
<p><i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i></p>	<p>Understands that it could be used by Library Services, Sector advocates, publishers, authors. potentially academics - but if it is the public domain - have you considered speaking to members of the public (patrons). Justification of the cost (public \$s) and the cost benefit. Would they use it or can they get the information they need from the public library catalogue and compare to other public library catalogues for example?</p>

Stakeholder 10	
<p><i>What is your understanding of open data and open data principles?</i></p>	<p>Good idea of open datasets. Has worked in international projects working with linked open datasets with BL and BNF in France. First thing we need to know is why – why this data, why is this useful. Feels like we're making assumptions and leaps about "why". The point of open data is about being open, it's about publicly releasing information that can be of use to anybody trying to answer a question.</p>

<i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i>	So what? Does knowing this stuff actually invoke an action? Or is it just "nice to know"? The dataset we have needs to be drawn back to a clear "why" otherwise there isn't much of a clear direction or reason. There are also companies that do this – why would we do it and can we even do it better? What's interesting about open data is how it's useful beyond libraries.
<i>If Yes, why?</i>	-
<i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i>	Doesn't see that there is much advocacy to do – is cheap, is easy, what's the need to advocate? Do have to do risk assessments in terms of the data you're making available.
<i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i>	Open data projects in other organisations (https://www.bl.uk/collection-metadata/downloads) tend to be pet projects from tech specialists, important to have a champion in a leadership space to say "this is a really good idea" and translate the importance to the rest of the organisation, how it can be an enabler or a tool rather than how it may be an end in itself. Important not to expect everybody to understand open data. Understanding the bridge between project and outcome. Understanding risk and knowing what your risk appetite is – where would you leap, where would you not leap.

Stakeholder 11	
<p><i>What is your understanding of open data and open data principles?</i></p>	<p>Broad understanding of open data from time working in university sector.</p> <p>Empathise with the challenge of data collation across public libraries in Victoria and highlighted the Libraries SA model that utilises a single LMS and membership model that grants access to all public libraries across the state.</p> <p>One example of the use of open data in a real world setting cited was the example of an open data/large data set being used to extend the life of a gold mine in Canada.</p>
<p><i>Do you feel there is a benefit to enabling public access to library collection usage statistics?</i></p>	<p>Yes. Very supportive of the use and availability of data.</p> <p>Highlighted that for the data to be of value it needed to be of a sufficient size. This will not only increase the validity of conclusion drawn from data, but also make the dataset more attractive and useful to researchers.</p> <p>Referenced Jim O'Donnell from Arizona State University and discussion on the value of data collection that informs decision making or outcomes . Is there a way that this data and its collection is of practical value to the every day, and informs decision making?</p>
<p><i>If Yes, why?</i></p>	<p>-</p>
<p><i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i></p>	<p>Highlighted the importance of presentation of data if it is to aid the sector and decision making.</p>

<p><i>What information do you require, or are you missing to make you more comfortable with the concept of library collection usage statistics open datasets?</i></p>	-
<p><i>In your opinion, who else do you feel would need to be consulted in order for this concept to be widely understood and accepted / would have considerable investment in this project being successful?</i></p>	-