A sea of 'parliamentary big data': collaborative approaches to its navigation by the Australian Parliamentary Library Research Service

Presented by Catherine Lorimer, Senior Law Librarian, Law and Bills Digest Section, Australian Parliamentary Library, Department of Parliamentary Services

The Australian parliamentary environment may be described as a microcosm of the big data world in which it operates. As part of its legislative process, policy development and inquiries, the Australian Parliament generates a vast number of documents in a variety of formats. In addition, further information providing commentary and analysis is published by the Parliamentary Library, media, academics and social media contributors. All of this information could be collectively described as 'parliamentary big data'. However, unlike conventional 'big data' it is a mix of qualitative and quantitative information; fact and opinion; truth and fabrication.

Like our parliamentary counterparts, one of the roles of the Australian Parliamentary Library research service is to navigate this data for clients—a task that is changing and becoming more complex. This paper will examine this challenge and how we are addressing it through the innovative use of technology and collaborative networks. The researchers require highly developed subject expertise, knowledge of parliamentary procedures and experience to identify and analyse this data in a timely and efficient manner. By developing and expanding collaborative approaches to research, information silos are being broken down and meaningful linkages between data sources identified. These linkages will be enhanced in the future as data mining techniques and visual data analytic tools are further developed and adopted.

What is big data?

Big data is defined as a collection of data too complex, too varied and too large for conventional database management tools to handle. Indeed, the classic definition of big data from the McKinsey Global Institute refers to big data as 'datasets whose size is beyond the ability of typical databases software tools to capture, store, manage and analyse'.

Big data remains difficult to understand as it can mean so many things to different people. To some extent it is a buzz phrase–one which is fluid, evolving and shifting with use and wider adoption. Essentially big data refers to two major phenomena:

- The breathtaking speed at which data is being generated every day
- The improving ability to store, process and analyse that data

Big data is not only about the scale or volume of data but also the scale of inter-connectedness, that is, the relationships or linkages that exist between large and sometimes disparate data sets.

The four V's of big data

According to IBM data scientists for data to be considered big, it needs to meet four dimensions – known as the four V's of big data:

- Volume–the sheer quantity of data. IBM estimates that 2.5 quintillion bytes of data are created each day. Of this data, 75 % is unstructured, meaning that it comes from sources such as text, voice and video rather than the more familiar structured data that is held in traditional databases. According to the book *Internet trends 2014*, it's been estimated that only 34% of this data it is useful; however only 7% of data has been tagged to give it any context and meaning, and only 1% of all data that exists has actually been analysed.

- Velocity–this is data being produced in real time, happening now, constantly growing and changing.
- Variety–data can come in different shapes, sizes and formats. Not just structured data but comments, blog posts, photographs, YouTube videos and tweets.
- Veracity–data needs to be tested for authenticity or trustworthiness. It needs to be reliable for meaningful conclusions to be drawn. If data-driven and evidence-based decision making is to be used in the development of public policy and for the delivery of public services, it needs to rely on high quality and accurate data.

Some data scientists have suggested another important V–value–how the data can be used. This is especially relevant to the role of the Parliamentary Library research service.

Role of libraries and big data
Librarians and researchers have been using big data for years. While the terminology may have changed over time, the role of libraries has been to collect, evaluate, and disseminate this data or information. We are expert users of structured databases and can apply these skills and expertise to new formats and volumes of data.
More recently the role of research services has been to value-add to this data; by efficiently identifying relevant sources, analysing them and drawing meaningful conclusions.

Parliamentary big data
In many ways the Australian Parliament is a microcosm of the big data environment world in which it operates. The Parliament generates its own 'parliamentary big data' and further data is generated about that data. As a member of the research branch, one of my roles is to use this parliamentary data and assist library clients in navigating this ever expanding sea of data.

What is 'parliamentary big data?'
As part of the legislative process and policy development, the Parliament debates Bills and policy in the parliamentary chambers; passes Acts and regulations; establishes committees and conducts inquiries.

Each of these parliamentary activities generates its own 'big data'. For example, the record of parliamentary debates in Hansard, votes, notice papers, media releases, committee hearings and reports, submissions and so on which are published in print or electronically. The audio and video records of parliamentary proceedings also contribute to this parliamentary big data. There are also numerous procedural documents which guide these proceedings.

The Australian Parliamentary Library also contributes to this 'parliamentary big data'. For example, by producing its own publications, developing collections, undertaking digitisation projects of parliamentary materials, writing individual client memos and sending out tweets about its activities. It also collects relevant press and academic articles, media programs which report and comment on the workings of the Parliament, politicians, political parties and policies, and adds them to structured databases.

In addition, there is hidden or private 'parliamentary big data' such as private correspondence, emails and social media comments; cabinet documents; internal party documents; departmental discussions and minutes.

This parliamentary subset of big data is unique in that it is a mix of qualitative and quantitative information; fact and opinion; truth and fabrication.

Impact of parliamentary big data on Parliamentary library clients
Our Parliamentary library clients–Senators, Members, parliamentary committees and parliamentary departments are aware of all this data being 'out there' and at times can feel swamped by it. They often refer to it as 'information overload'.
In fact the first mention of the phrase 'information overload' in the Commonwealth Hansard occurred in 1979 on the occasion of the tabling of the Department of the Parliamentary Library annual report. The Speaker at the time noted:

> Honourable members know only too well the amount of information to which they are subjected to in the course of their parliamentary duties. While this seems excessive it is only symbolic of the tremendous amount of data now available and of the tasks of official information services in selecting, storing, retrieving and delivering it as, when and where required. In recent times there has also been a great increase in the complexity of current issues and in the speed with which information must be assimilated and disseminated. These changes mean that library and information services such as those of this Parliament are faced with demands for digests, summaries, analyses and objective selection to provide relief from 'information overload ' and give essential information in the most concise and relevant form.

These comments were made in the era when most information was provided in paper form only. Fast forward to the 21st century where there is now a veritable tsunami of information in a variety of formats, a 24/7 news cycle and social media commentary.

Role of Parliamentary Research Service in navigating Parliamentary big data
Expertise of staff
The research service at the Australian Parliamentary Library is comprised of seven subject areas: economics; foreign affairs, defence and security; law and bills digest; politics and public administration, science, technology, environment and resources; social policy and statistics and mapping. There are currently 75 researchers working across these subject areas.

The research service assists our clients to navigate this sea of parliamentary data by guiding them to, and providing them with the information they require through the innovative use of technology, as well as utilising human networks. While our remit is to use only publicly available information, the experienced librarians and researchers have the knowledge and expertise in finding, navigating and analysing these documents.

On receiving research requests, researchers will apply their own subject expertise and knowledge in providing a response. However, this expertise also needs to be coupled with a working knowledge of parliamentary procedures and documents, so a parliamentary perspective can be provided if required.

For example, a request on euthanasia may require the legal and social arguments for and against, but also information on whether the Parliament has previously considered this issue; whether there been Bills introduced, was a conscience vote allowed, what was the final vote, how did certain individuals vote, was there a committee inquiry into the issue, and how many public submissions were received.

Technology and search expertise
The most recent library client survey indicates that about 75% of Senators, Members and their staff use Google as their first place to find information. However, when this does not provide what is required quickly they will often then contact the Library for assistance in locating what they need. They also view the Library as a trusted and reliable source of information.

Being at the frontline of receiving and responding to requests it is often clear to me that clients know that the information they require is available but they are unsure where to locate it and how search for it, especially when they need it quickly.

To continue the nautical analogy, they are 'drowning' in a sea of parliamentary data and need help in navigating it. I have often received phone calls from staffers with the opening line 'I have searched on Google but can't find what I need' or 'I should know where to find this but am not sure where to look'. Often at this stage, the required information is needed quickly, as it may be required for a media comment or to respond to time critical issues such as a natural disaster. For example, information on flood insurance and disaster relief payments was required urgently following Cyclone Debbie in Queensland and New South Wales.
This is where the expertise of the research librarian is applied to retrieve and deliver the requested information. We have the skills to target relevant and appropriate information sources which is more efficient that data trawling across Google.

In this current era of fake news and alternative facts, there have been requests to 'fact check' particular news stories and statistics; about 20 requests so far in 2017 have used that specific phrase. Again the expertise of the research librarian in determining the reliability and credibility of these news stories is invaluable but it must be noted that this is not a new phenomenon. Keys to checking the veracity of information source are its currency; its relevancy; its authority; its accuracy and its purpose.

Parlinfo and the APH website
In the Australian Parliament, much of the parliamentary big data is available through the Parlinfo database. While the data is arranged into discrete databases, it is possible to search across all this data at once or to only search selected databases. There is also a wealth of parliamentary big data available through the Australian Parliament House website especially in relation to committee inquiries.

In addition to daily uploads of news, articles, press releases, and media, there are a number of specific projects currently underway to digitise parliamentary documents. These digitising projects include all Parliamentary papers published since 1901 to 2012; all Bills and

associated documents since 1901 and the conversion of the analogue audio and video recordings of parliamentary proceedings to digital formats.

However, as these large volumes of historical parliamentary data are added to Parlinfo, the result sets get larger and need to be refined to give meaningful answers to clients. This can be done by using facets such as date, source and relevance.

Limitations of the technology

However, the current Parlinfo search system and APH website search criteria capability both have limitations.

For example, public submissions received by parliamentary committees are available as discrete documents on the relevant inquiry website and can be selected and downloaded individually. Over time, the volume of public submissions to committees has increased, especially on controversial topics. Some of these submissions provide very valuable data on an issue but unless referenced in the final committee report are not readily known.

There is currently no search functionality that allows searching across all submissions to an inquiry. This would be useful where a large number of submissions are received to discover common themes or where specific legislation is mentioned. The Parliamentary departments are working collaboratively to improve the design and usability of this search function.

Law and Bills Digest Section

The Law and Bills Digest section uses parliamentary big data sources extensively. As well as preparing and publishing Bills digests on all Government Bills, individual client research requests are undertaken.

These requests can range from the relatively simple and straight forward such as locating a current version of an Act to tracing the legislative history and debate of a particular provision of legislation.

For example, a recent request was to determine the number and volume of legislation passed by the Commonwealth from 1901. Finding the number of Acts and regulations passed was relatively straightforward and was compiled using a number of data sources including old bound volumes and different legislation databases.

However, the parliamentary data did not provide the total number of pages of legislation passed. This was because the cumulative pagination of Acts ceased in 2001 and that data was not readily available.

More complex requests can require comparisons of laws between international jurisdictions. For example, a request for comparative domestic violence laws requires access to parliamentary big data from other legislatures in order to provide an analysis of the different approaches adopted on this issue, and any subsequent reviews of their implementation.

Collaboration

Collaboration is essential to navigating this sea of parliamentary data. This can encompass utilising technology and human networks.

- *Collaboration within the research branch*

Since 2010, all research branch client memos are required to be captured into TRIM which is an electronic document and record management system. Prior to this, each section stored their client memos on work drives which were only accessible by that section.

With the introduction of TRIM, each section is responsible for filing its own memos and entering specified metadata for each document. The great advantage of using TRIM is that client work can now be searched by all library researchers. This means that previous work on a given topic can be identified. This has started the process of breaking down the information silos that have previously existed between sections and also provides workplace efficiency.

Access to client work on TRIM has led to greater cross-collaboration between different sections as researchers can identify who has worked on that subject previously and encourages interaction with them. One advantage of this process is that where new developments have occurred in that subject area, a conversation about the currency of previous work can occur and ensures that outdated information is not provided.

Of course, the issue of client confidentiality is paramount in any document management system. To this end, access restrictions are in place for client work. Subject sections can see the full-text of their own client memos but they can only see the metadata of client memos from other sections and then request to view the full-text.

- *Collaboration with Parliamentary Committees*

The 2015 client service evaluation identified that the Parliamentary Library needed to be more pro-active in engaging with parliamentary committees. In the past year, steps to build relationships with committee secretariats and research staff have been undertaken.

This has seen the Library contact secretariats when new inquiries are announced and discuss how the research branch can assist with providing background information to the committee.

Additionally, there have been opportunities for library staff to be seconded to committees and vice versa. This has led to a greater understanding of the procedures, processes and focus of the work undertaken in these environments.

- *Collaboration with State and Territory Parliamentary Libraries*

As Australia is a federal system, there are many areas of law where state and territory laws apply rather than one national law, for example surrogacy laws.

We regularly assist our colleagues in the state and territory parliamentary libraries, and we seek assistance from them, especially when a response is required quickly and this is best obtained through personal contacts in the research service.

For example, the proposed establishment of a parliamentary injury compensation scheme for federal politicians required investigation into what type of schemes, if any, operated in the Australian states and territories as well as overseas. Emails sent to our state and

territory counterparts seeking assistance provided a wealth of information to respond to this request.

- *Collaboration with International Parliamentary Libraries*

When information is not readily found via the websites of overseas legislatures, email requests to overseas colleagues are used. For example: with the injury compensation scheme, emails were sent to international colleagues and useful responses supplied.

This year a joint project with research colleagues from overseas parliaments has commenced. A research paper on the legal framework regulating intelligence agencies is being undertaken with contributions from research colleagues from Australia, Canada, New Zealand, the United Kingdom and the United States resulting in a final collaborative paper. This is a good example of parliamentary research services working together and further developing personal contacts and networks.

- *Collaboration with external experts*

Our collaboration with external experts acknowledges the fact that the library research staff cannot hope to maintain expertise across all areas of specialist knowledge. This collaboration happens to some extent through the Library topical seminars presented by subject experts, short term contracts to prepare specialist papers, external peer reviews of publications and more recently, specialist briefings. These briefings are limited to library clients, where questions can be asked following the expert providing an oral briefing. A specialist briefing was recently held on India, where three MPs and 12 political staff attended.

Future directions

The concepts of 'data mining' and 'data analytics' relate to powerful new computer processing techniques which are used to discover, process and analyse vast, inter-linked data sets to identify patterns, trends and sentiment. These tools have enormous commercial and public sector potential.

The Parliamentary Library recently introduced access to Buzznumbers which monitor trends and sentiments in social media, such as Twitter and public Facebook pages, against a customised profile. For example, a profiles that monitors tweets that mention the Prime Minister, the Budget or a particular government agency.

The Parliamentary Library's 2015 *Future vision report* identified that collaboration will become increasingly important in order to respond to complex questions. Data analysis skills will also need to be further developed among researchers in order to enable them to consider and understand the potential of data analysis, data sources and interpret the results.

Concluding comments

Looking to the future, it is clear that the sea of parliamentary data will continue to expand in a myriad of formats. Collecting this data, investigating it, and analysing its depths and inter-connectedness present continuing challenges for library researchers. However, the analytical mindset of librarians and researchers is well-suited to identify and collect valuable information from this great sea of big data.

While technological developments in data mining software and visual data analytics will provide some of the answers to discovering important linkages and synergies that exist in parliamentary big data; the fostering of human networks and collaboration both within the Parliament and with external experts is equally important.

References

Australian Parliamentary Library, *Future vision report*, July 2015.

Bates, M, Big data ain't so big, *Online searcher*, 41(1), 2017, p. 72.

Cato, A, Big data? Big deal?, *Australian Law Librarian*, 22(4), 2014, pp. 222-224.

Fishleigh, J, A non-technical journey into the world of big data: an introduction, *Legal Information Management*, 14 (2), 2014, pp. 149-151.

Gordon-Murnane, L, Big data: A big opportunity for librarians, *Online*, 36(5), 2012, pp. 30-34.

Missingham, R, Parliamentary library and research services in the 21st century: a Delphi study, *IFLA Journal*, 37(1), 2011, pp. 52–61.

Uncommon Knowledge, *Australian Parliamentary Library: Client service evaluation*, 2015.

Williamson, A, Big data and the implications for government, *Legal Information Management*, 14 (4), 2014, pp. 253-257.