



contentgroup

TRANSCRIPT OF PODCAST

Work with Purpose

EPISODE #68

REALISING THE POTENTIAL OF DATA IN GOVERNMENT

Hosted by Cheryl-anne Moy, Deputy Secretary, Deputy Secretary of Immigration and Settlement Services, Department of Home Affairs, and IPAA ACT Deputy President

TRANSCRIPT

David Pembroke:

Hello everyone and welcome once again to Work with Purpose, a podcast about the Australian Public Service. My name is David Pembroke. Thanks for joining me. Today, Work with Purpose takes you back to an event held in June of this year when Dr David Gruen, the Australian statistician, presented on the topic of realising the potential of data in government. Dr Gruen builds on an earlier presentation, the promise of data in government and emphasises that the pandemic has certainly accelerated the use of data to meet public policy challenges. He explores the critical role of data in responding to the pandemic and the importance for future collaboration in deriving public value from data. Dr Gruen's address was followed by a conversation with Cheryl-anne Moy, who is Deputy Secretary of Immigration and Settlement Services at the Department of Home Affairs and she is also an IPAA ACT board member and deputy president. The episode begins with the voice of IPAA CEO, Caroline Walsh.

Caroline Walsh:

Welcome and thank you for joining us today. I'd like to start by paying my respects to the Ngunnawal people, the traditional owners of the lands on which we're meeting today and extend my respect to their elders and thank them for their ongoing care and custodianship of the land. Welcome to our event today, realising the potential of data in government. Thank you for coming along to another IPAA event and thank you to Dr Gruen for his support of IPAA and using IPAA as a platform to support, debate and striving for excellence in public administration. We're really delighted to be able to bring events like this to you that support your professional development and your networking across the public sector.

Our IPAA host for today is Cheryl-anne Moy, who is a deputy secretary at the Department of Home Affairs. She's also an IPAA ACT counsellor, IPAA ACT board member and the chair of the IPAA ACT Corporate Governance Committee. She has a few roles to play. I'd now like to welcome Cheryl-anne to the stage. Thank you.

Cheryl-anne Moy:

Good morning, everybody. It's great to see everyone out here on a cold winter's day and I think we've got a few more of them to come. I'd really like to acknowledge all the senior executives, guest, members and sponsors that are here today and I echo Caroline's acknowledgement of country of the Ngunnawal people. It's my pleasure to introduce our keynote speaker today, Dr David Gruen AO, who was appointed Australian Statistician on 11 December 2019. David was previously the Deputy Secretary Economic in Australia's G20 Sherpa at the Department of Prime Minister and Cabinet. Before joining the Department in September 2014, he was the executive director of the Macroeconomic Group at the Australian Treasury. David joined the treasury in January 2003 before which he was the head of Economic Research Department at the Reserve Bank of Australia from 1998 to 2002. David holds PhD degrees in physiology from Cambridge University in England and in economics from the Australian National University. David was appointed an officer of the Order of Australia General Division in 2022, congratulations, for distinguished service to public administration, economic research, business and education. Please help me welcome Dr David Gruen to the stage. Thank you.

Dr David Gruen:

Thanks very much. Today I will describe the transition from the potential of data in government to realising that potential. In the time available, I will limit myself to three broad topics. Firstly, I will talk about some of the new data sources that have become available and how they are being used to generate new statistics. Second, I will talk about the growth in integrated data assets across many dimensions. These dimensions

include improved timeliness, the number of integrated data assets, the increasing variety of the subject matter they cover and the strong growth in usage both by government agencies and researchers. Third, I will reflect on the opportunities presented by the passage of the Data Availability and Transparency Act. When I addressed IPAA in March 2020, the COVID 19 pandemic was in its infancy, but it was already looking like data. Both health data and data about the economy would play a pivotal role in helping guide the government policy makers and the community through the coming challenges.

Elsewhere, I have spoken about the early ABS responses to the pandemic, introducing new small rapid business and household surveys and publishing a range of preliminary and provisional data. I will not go over this ground again here. I do want to talk about the new data sources we are accessing to provide statistical information that was previously unavailable. Data from these sources are all byproducts of the digital revolution, a consequence of so many aspects of our lives now being intermediated through digital platforms and another general point, all these new data sets are examples of big data, large and usually complex data sets from new sources. Let me begin with single touch payroll. The Australian Tax Office receives payroll information from employers with single touch payroll enabled payroll software each time the employer runs their payroll. Given the extensive coverage of the STP system, these data cover more than 10 million employees.

That is not quite every employee in the country, but it's not far from it. The pandemic made access to this rich vein of near realtime information an urgent priority. The ATO expedited access and the ABS began receiving these data in early April 2020. From then on, each week, the ATO provides job and wage data from the single touch payroll system to the ABS with which we produce a new publication, Weekly Payroll Jobs and Wages. We are particularly imaginative when it comes to our titles. In many ways, access to single touch payroll data taught us new ways of doing things. Given the scale and complexity of these data, it made sense to ingest and analyse them using cloud computing services rather than our existing computer systems and that is the new model for accessing public and private sector big data assets to generate new statistical insights.

Let me describe a few of them. In October 2021, we began releasing a new monthly indicator of business turnover based on business activity statements submitted to the ATO. Again, to give you a sense of scale, there are about 130,000 remitters of business activity statements from whom we gather information for this new monthly indicator. The 130,000 should be compared to our comparable survey, the quarterly business indicators, which is based on a sample of 16,000 businesses. In February this year, we released a second monthly indicator, which provides a measure of household consumption. This indicator is based on about 800 million bank transactions by households each month with those data provided by Australia's major banks in aggregated de-identified form. Household consumption accounts for about half of GDP, so there's considerable value in having an accurate measure of it. The existing monthly measure of household consumption comes from the Retail Trade Survey based on a sample of around 3,400 businesses.

The Retail Trade Survey covers about 30 percent of household consumption. It used to cover more, but households keep moving their consumption away from the things we measure in the Retail Trade Survey, which is inconvenient of them. Thirty percent of household consumption is picked up by the Retail Trade Survey, whereas the new measure based on banks transaction data covers about 68 percent of household consumption so that's a substantial step up. Many items of household consumption are captured in the new transactions based measure, but are missing from the Retail Trade Survey. To give you a few examples, a particularly topical one, the purchase of petrol, car servicing and maintenance, train and bus tickets, Uber rides, airline tickets, hotels, theme parks, haircuts, dentists and allied health costs, all things that people pay for with their credit cards. None of these are in the Retail Trade Survey, but they're all captured in the new indicator.

We're also developing a partial monthly indicator of the Consumer Price Index, which is particularly relevant given the current inflationary environment. This has been aided by our access to digital data sources, including scanner data from supermarkets and web scraped price data. We plan to release an information paper within

the next couple of months and begin publication of the partial monthly indicator later in the year. We will also be releasing a new monthly indicator of individual earnings in 2023, again using single touch payroll data.

A significant benefit of using existing data collected for other purposes to generate these new indicators is that there's no need to put a new survey in the field, which places an unavoidable burden on respondents to that survey. The digital revolution also offers new ways to reduce the burden on our existing survey respondents. We are working with businesses, accountants, bookkeepers and accounting software companies to co-design a new reporting application that links with the accounting software that businesses currently use. In the future, a business will have the option to extract and pre-fill their financial data directly into an ABS web application from their accounting software package. This removes the need for businesses to manually collate the information and key it into our surveys. Once this is up and running, we estimate there will be a 16,000 hours per annum saving or 70 percent less time that small and medium businesses spend completing ABS surveys. As part of the new initiative, we will provide tailored reports back to business to help them understand their performance relative to their peers.

When I spoke at the IPAA event in March 2020, I described the growing number of integrated data assets being used across the public sector to enable research policy development and analysis. I focused on BLADE, which is the Business Longitudinal Analysis Data Environment and MADIP, which is the Multi-Agency Data Integration Project and I need to meet the people who came up with those acronyms, which are the business and person integrated data assets, which have been developed and enhanced over many years by the collaborative efforts of many people across many commonwealth agencies and departments. In the little over two years since that talk, there has been a lot of progress moving from the promise of integrated data assets to realising the benefits they can provide in the service of better public policy. Let me describe some of this progress. Our earlier standard practice was to update the underlying data in both BLADE and MADIP once a year, but as these data assets have matured, processes have been streamlined and key enabling infrastructure has been moved to the cloud.

This enhances security and makes possible more sophisticated data analysis, but it also means that BLADE and MADIP can now be updated much more frequently. I'm going to describe some of the much more frequent updating. There have been many additions to these integrated data assets and I'll tell you about a few of them. We've introduced a new quarterly updated business locations data set to BLADE, so that's the business integrated data asset, which enables detailed geospatial economic analysis. This addition to BLADE with quarterly updates of BAS data recently allowed the ABS to provide the National Recovery and Resilience Agency with detailed geographic business counts and economic information for the flood devastated areas of New South Wales and Queensland. If you have detailed geographical information, you can tell the relevant people where the businesses are and whether they're at risk of being flooded.

Single touch payroll data are being integrated into the core of the BLADE data asset and will be available to approved researchers within weeks. The core data in MADIP and BLADE have now been linked together and as a result, researchers have been able to use these two assets to undertake research and analysis of linked employer and employee longitudinal information to determine the impact of COVID 19 on businesses and people and to examine economic recovery and employment and unemployment patterns. Data from the Australian Immunisation Register are being linked to MADIP each week. I don't think the people who were in this project thought they could do this each week till someone said, "How about we do this each week?"

Provisional death registration data are being linked and updated monthly. These data are being used by the Department of Health to generate insights for the Australian COVID 19 vaccine and treatment strategy and by state health departments and primary health networks. Here are a few specific examples. The National COVID 19 Vaccine Task Force identified groups across Australia with low vaccine uptake who spoke languages other than English. In response to this information, culturally appropriate communication campaigns, digital translations and community outreach activities were implemented to lift vaccine rates for these groups and you can see why it's critical for it to be up to date. It wouldn't be much used if it was updated a year ago.

Jurisdictions have deployed multilingual GPs and healthcare workers to better support multicultural communities. The provisional deaths registrations data are being used alongside a range of other sociodemographic information to understand risks to vulnerable groups within the community, as well as to support winter preparedness strategies. These data have also been used by peak technical advisor groups like the Australian Technical Advisory Group on Immunisation, the well known ATAGI and the Australian Health Protection Principles Committee to inform their decision making. To support treasury analysis, the Labour Market Tracker Project, integrated job related data, including single touch payroll, job keeper and job seeker data to both BLADE and MADIP. These data sets are updated fortnightly, monthly and quarterly as they become available to enable up to date monitoring of the labour market and the economy.

A further key data integration project is the National Disability Data Asset. The NDDA is under development and will include a collection of linked de-identified data sets from across multiple Commonwealth, State and Territory agencies to better understand the lives of people with disability and their pathways through services. The NDDA will be underpinned by a new national data integration infrastructure known as the Australian National Data Integration Infrastructure. The ABS, the Australian Institute of Health and Welfare and the Department of Social Services are partnering to deliver this initiative. The NDDA and the data infrastructure will be co-designed and co-governed with State and Territory partners as well as the disability sector. The integration infrastructure, it's called ANDII, Australian National Data Integration Infrastructure is being built in such a way that it can be reused in public policy domains beyond disability.

My final example gives you a sense of the breadth of subject matter areas in which data integration projects are being developed. It goes by the name, the Justice Spine. You could think of all sorts of things, but that's what it's called. It is a longitudinal national data asset linking police recorded criminal offenders in Australia's criminal courts with adult prisoners in the corrective services system in every State and Territory in the country. That's a lot of different data sets from a lot of different entities. The data set will show how people move and interact within and across the justice system nationally, something that is currently not possible. Clearly, people can move from one State to another, and it's important that you can track that movement. That's why you need a national system. The data set will have the potential to be linked to other Commonwealth, State and Territory held data sets for deeper analysis of the characteristics of criminal offenders. It will be available to approved policy makers and researchers in late 2023 and will enable analysis of patterns of offending and policies to reduce recidivism.

There is little point sharing and integrating data if it cannot be accessed. Increased data integration capability is being complimented by a data access service, the ABS Data Lab. The ABS Data Lab enables sophisticated analysis of detailed micro data in a secure controlled environment. Use of the ABS Data Lab is currently growing at a compound rate of about 30 percent a year. To give a sense of this growth, there were about 50 data lab users in 2016, almost 900 by 2019 and there are now around 4,000. There are 400 active projects currently on the go across governments, both Commonwealth and State and Australia's research sector. About 60 percent of them are government projects and about 40 percent are academic projects either from universities or from research institutes. The data lab is also being made available as a platform for data sharing. The Department of Finance and the Australian Tax Office are tooling up to use the ABS Data Lab to enable secure sharing and sophisticated analysis of their data.

The richness of data sets now available in the ABS Data Lab has significant value for academic research. With appropriate safeguards in place, we are now piloting access for international academic researchers by partnering in the first instance with the OECD and with Professor Greg Kaplan at the University of Chicago. By providing international access to what are now high quality data assets with appropriate safeguards, there is the prospect that more international researchers will be attracted to working on Australian policy issues using Australian data. This can only help in generating new insights on Australia's policy challenges. In the academic discipline I know best, economics, it's hard for academics to get research on Australian economic issues published in the top international journals. Making Australian data available to international researchers should generate more interest in Australian policy issues that can be tackled using these data. In turn, this

should make a modest contribution to improving international recognition of academic work conducted using Australian data.

Let me turn to the Data Availability and Transparency Act. On the 31st of March 2017, the Productivity Commission sent its Data Availability and Use Inquiry Report to the Treasurer, the Honourable Scott Morrison MP. The commissioners on the inquiry were the chair of the Productivity Commission, Peter Harris and Melinda Solento, now the chief executive of the Committee for Economic Development of Australia known as CEDA. Five years later to the day on the 31st of March 2022, the Data Availability and Transparency Act, that's not an accident, that spells out DATA, received Royal Assent. The Act establishes a new best practice scheme for sharing Australian government data underpinned by strong safeguards. The data scheme is focused on increasing the use of Australian government data to help deliver government services, inform government policies and programmes and support research. It will provide strong support for better government data use and collaboration. Importantly, the DAT Act enables the sharing of Commonwealth data with the States and Territories as well as with academics.

It is worth explaining the relevance of the DAT Act for the ABS. The ABS' legislation supports making aggregate data publicly available and the ABS Data Lab enables authorised access to detailed micro data while protecting the privacy of individuals. Increased data sharing under the DAT Act will increase the need for secure sharing and access infrastructure and there are opportunities to expand the use of the ABS Data Lab as a service, so agencies like the ATO can use the DAT Act to share data safely to a range of users. There are also opportunities for us to use the DAT Act to streamline our data sharing with other agencies, including by making it easier to share de-identified data with trusted partners such as the Australian Institute of Health and Welfare. Agencies across the Commonwealth will now have a streamlined legal pathway to share the data with the ABS like the Department of Education Skills and Employment sharing apprentices and traineeship data for data integration projects.

Let me wrap up. It is instructive to look back at the recommendations on the use of data from the 2019 independent review of the Australian Public Service, the Thodey review. The review made a strong case for enhancing the use of data to support public policy formation and better service delivery. Many of the things I've talked about today were foreshadowed in the review as important directions for the future of the Australian Public Service. Among these are accessing new data sources for public policy purposes, wider use of integrated data assets to rigorously develop and improve policies and legislation and infrastructure to enable data to flow securely between agencies. The Thodey review also recommended that the public service launch linked data and digital professions to build data and digital expertise, which of course has also come to pass. There is always more to do and we can't rest on our laurels, but equally, there has been impressive progress over the past few years in helping to realise the potential of data in government. Thank you.

Cheryl-anne Moy:

Dr Gruen, thank you for that. That was very informative and following on from your speech of 11 March, which we really did just manage to get in before the lockdown started all over Australia, but particularly Canberra for us. In terms of that speech and the promise of data in government and your lead on today, there's been a lot of work and I don't think that you had so much problem with the acronym for the DAT Bill or the DAT Act, I think that one was okay. What do you think have been the challenges for ABS in working collaboratively with agencies around data in not only the COVID period, but in the immediate period especially around data usage? I know you talked, for example about some data being used during the COVID period to assist with decisions that were made and day to day decisions around vaccinations, et cetera, which, as an agency that worked very heavily in that, I'm very grateful. What do you find are the challenges though in dealing with agencies around data usage?

Dr David Gruen:

I always feel uncomfortable making this point, but I think COVID was a wonderful thing, clearly not in the broad, but for sharing across the public service and particularly data. Obviously the health data played an extraordinarily critical role. We were getting updates, a whole lot of new people became household names, and plenty of that was based on the data that they were talking about, but it was also an opportunity for breaking through some of the natural hesitancy about sharing data. I saw plenty examples of that. Situations where you thought it was going to take months to get access to something and people understood the urgency and were willing to accommodate it really quickly and that was not just the public sector, it was the private sector. One of the things that we did and I think quite a lot of agencies did this, the banks were extremely cooperative, so they didn't charge.

They just said, "How can we help?" We got access to transactions data from the major banks. I don't know exactly when, but I wrote to them around March or April of 2020 and they all wrote back and said, "We can help," and I know that several other departments like Treasury got access to this as well. I think the other thing about COVID which and I've made this point before, in contrast to the global financial crisis where you just didn't have all these sources of real time data, on the nightly news you were getting, there was that wonderful data set for restaurant bookings around the world, which you'd think was kind of weird, but in fact it was quite informative about whether people were going out or not. Data was front and centre for a lot of the community.

The other thing we talk about a lot is the decline in trust in government and that turned around. Maybe it's not permanent, but the public saw the things that the public sector could do for them and measures of trust went up and that's encouraging. My answer is that COVID was an opportunity for more data sharing across the system and a general sense that that was the right thing to do and maybe now that, it's not post COVID unfortunately, but we've moved to a stage where we're actually meeting in rooms with large numbers of people again, but the hope is that we keep all those gains.

Cheryl-anne Moy:

The ability to maintain those connections outside of an emergency zone is and would be very beneficial. The single touch payroll that you talked about was a really good example of ATOs engagement with industry. How do you think other departments and agencies or as an APS, we can continue to expand beyond different sectors and boundaries and jurisdictions and still maintain that trust that people have in government around the use of data. How can we expand in similar ways across departments and agencies say in a way that ATO did with the single touch?

Dr David Gruen:

The ATO does have the advantage of compulsion, which is...

Cheryl-anne Moy:

Always. Always. They actually compelled the public service. I was procured at the time. I remember it well.

Dr David Gruen:

Yes. They have that benefit, but also the digital revolution provides the opportunity for making so many things easier and the public is on board mostly with the idea that if you can make my life easier, then I'm all for it. Obviously, there's been an enormous amount of digitisation of things and so I think lots of departments are involved in trying to streamline the service delivery that they provide to the community. Part of that is making it more digital and almost always on the back of a digital platform is a whole lot of data and so the idea is, use that data, use it wisely, make sure that you have appropriate safeguards, but almost certainly that data is useful for something.

Cheryl-anne Moy:

Speaking of which, the recent census, ABS won the spirit of service award from IPAA.

Dr David Gruen:

It did.

Cheryl-anne Moy:

For that arrangement. Congratulations to them.

Dr David Gruen:

Us and ASD for the cyber protections on the census. Yes, indeed.

Cheryl-anne Moy:

How do you think the data collection from the census will mark a new way of collecting and storing and sharing and using data in the future?

Dr David Gruen:

Obviously after what happened in 2016, everybody who was working on the census was acutely aware that we had to make sure that that didn't happen again and we couldn't be completely certain because nothing is certain in the digital world, but there was an enormous effort put into making that a much more user friendly experience and safe and I think it was a credit to the public service that it went off without a hitch and that it was for most people a very enjoyable experience. I think the public sector is getting increasingly good at providing digital service to the community. I happened to have to create a COVID vaccination international certificate and I thought, "What's this going to be like?" That's incredibly straightforward. You go onto mygov and it takes two minutes and so we are increasingly doing this with our services, so I think it's one area that's got bipartisan agreement that improving service delivery for the community is something that is high on the agenda for the public service and I think lots and lots of departments are involved in it.

Cheryl-anne Moy:

Yeah, I agree. I think any policy framework now comes with a discussion first about what's the service delivery experience that we expect for clients at the end, which will go well into the future. Do you think there's a disconnect between our regulatory frameworks in terms of the storage and usage, the collection and whether or not that regulatory framework is keeping up with our use of data and what we need from data?

Dr David Gruen:

I think regulatory frameworks always have trouble keeping up with rapidly changing environments. I think it's probably right that the regulatory frameworks can do with a refresh. Having said that, there's often a lot you can do without necessarily changing the legislation. There are risks as well as opportunities with opening up legislation, but nevertheless, my favourite example of that is that the Reserve Bank of Australia has a inflation target, which it invented with the help of the government on the basis of its current legislation, which dates from 1958 when we had a fixed exchange rate. They just did it and they haven't updated the legislation because opening up the Reserve Bank Act, you just never know what someone will put into it. My answer is, you might be able to change the regulations without necessarily changing the legislation and that's often a way forward, but if you get the chance, then good to update legislation if you can.

Cheryl-anne Moy:

It's my current life with the Migration Act of 1958.

Dr David Gruen:

Right. Yeah.

Cheryl-anne Moy:

Post-war. Interesting. You mentioned a lot of new products that ABS has launched or about to launch. What do you think are the new frontiers for data? Where do you see things in five to 10 years?

Dr David Gruen:

There are two directions for data and they're very different. One is big data, which I talked about and integrated data. Those things are going to continue to grow and they're going to continue to generate sophisticated insights you can't generate with single individual data sets. That's one direction. The other direction is social surveys on sensitive topics. There's a lot of demand for high quality surveys on things like domestic violence, all that sort of thing, which requires sensitive face to face interviewing with people in order to get that information. In a sophisticated society like ours, there's a lot of demand for having that sort of information because understandably, where we are now to say with domestic violence is a place where there is room for enormous improvement and so that's a whole other area, which is, as I say, they're quite different things. That's not going to be affected. I suspect it's not going to be changed radically by the digital revolution. That's about social interactions between people, but information about that stuff is also extremely important.

Cheryl-anne Moy:

David, thank you so much for presenting today and sharing your insights with us on the challenges and opportunities of data. It certainly has been informative. I know that we've had a very engaged audience and we really appreciate your time and your wisdom and insight on those issues. Thank you again for your presentation today.

Dr David Gruen:

Thank you.

David Pembroke:

There we have it. Cheryl-anne Moy closing out that fascinating presentation and conversation with one of the great leaders of the APS, Dr David Gruen, who is always worth listening to. Work with Purpose is a part of the GovComms Podcast Network. A big thanks to our friends and colleagues at IPAA who have helped us to put together such great content today and also to the Australian Public Service Commission for their ongoing support. A big thanks also to the team at contentgroup who help us with the technical production of the program. We'll be back at the same time in a fortnight. My name is David Pembroke. Thanks for joining me and it's bye for now.

Voice Over:

Work with Purpose is a production of contentgroup in partnership with the Institute of Public Administration Australia and with the support of the Australian Public Service Commission.