Emily Sterling :

Hello, and welcome to this morning's first keynote address for... It's on a topic of a data-driven APS, Reflections and New Directions. My name is Emily Sterling, and I'm a member of the Graduate Data Network. It is my pleasure to introduce our keynote for today, Dr. David Gruen. Dr. Gruen was appointed the Australian Statistician of the Australian Bureau of Statistics in December 2019. He was previously the Deputy Secretary, Economic and Australia's G20 Sherpa at the Department of Prime Minister and Cabinet.

Before joining the department in September 2014, he was executive director of the Macro Economic Group at the Australian Treasury. Dr. Gruen joined the treasury in January 2003. Before which, he was the head of the Economic Research Department at the Reserve Bank of Australia from 1998 to 2002. Before joining the Reserve Bank, he worked as a research scientist in the Research School of Physical Sciences at the Australian National University, and he holds PhDs degrees in physiology from Cambridge University and economics from the Australian National University. Please join me in welcoming Dr. Gruen.

Dr. David Gruen :

Thanks very much, Emily. I will just have to get used to speaking to an empty room, but you can imagine that there's actually an audience in front of me, and you're just viewing the proceedings from afar. Let me begin by acknowledging the Ngunnawal people, the traditional custodians of the land on which I'm speaking, and pay my respects to their elders, past and present, and extend that respect to any Aboriginal and Torres Strait Islanders who are watching this broadcast. Thank you to IPAA and the Graduate Data Network for providing the opportunity for me to speak today, even though it has to be done virtually rather than face-to-face.

I'm going to talk about a few different things today. I'm going to talk about the opportunities that COVID-19 has presented for an organisation that spends its time producing statistics and data. I'm going to recap some of the lessons from the IPAA speech I gave on the 11th of March called The Promise of Data in Government. It feels like a long time ago now, the 11th of March. There was a room full of people in front of me at that time, but no longer. That's not the world we live in, at least not for some time. Then finally, I'm going to talk about some of the exciting new opportunities that that are opening up for data and data analysis, things that we're just beginning to do. So that's the order of my talk.

Just at the moment, it's hard to talk about anything other than COVID-19. In less than a month, the spread of COVID-19 has turned many of our lives upside down and led to rapid fire responses by governments, each more remarkable than the one before that none of us have seen in our lifetimes. Importantly, for the topic today, the spread of COVID-19 has also highlighted the importance of high-quality data in circumstances where it is particularly difficult for the community and policy makers to quickly and accurately assess what is going on.

With COVID-19 on my mind, let me begin by talking about some of the responses the ABS has in train to provide data that is as up-to-date as possible to inform the community and policy makers. In a press release I issued on the 16th of March, I said this, "The economic implications of the spread of the coronavirus, COVID-19, are highly uncertain. In these circumstances, there are sizable benefits for the community and governments to have access to information about the economic responses of individuals and businesses that is as up-to-date as possible. The ABS has considered what additional more up-to-date information it can provide over and above the existing statistical releases to enhance understanding of the economic impacts of the coronavirus." So that's the quote from the press release.

I then listed a series of extra surveys the ABS is running as well as other data we committed to release in preliminary form more quickly than normal. The first new survey we ran, Business Impacts of COVID-19,

was designed in double-quick time with help from Treasury, guiding us on the information they thought would be valuable to collect. Given that it was no longer safe to conduct face-to-face interviews, this survey was conducted exclusively over the phone. Interviews began on Monday, the 16th of March, and were completed on Monday, the 23rd of March. We had planned to collect data for the survey for longer than this, but stopped early because the Australian government announced stage one restrictions on social gatherings on Sunday, the 22nd of March. That was a profound change in the economic environment facing many firms which would have created a big structural break in these data had we continued to collect it.

Six days of interviews enabled us to collect responses from just over 1,200 businesses and provided enough detail for us to report results desegregated by business size, that is by number of employees, and by industry sector. Having finished collecting data for the survey on the 23rd of March, we published the results on our website on the 26th of March three days later. I can't be certain, but I suspect that that may be the fastest survey ever conducted from start to finish by the ABS. I don't know who came up with the phrase, "Never waste a crisis," but it seems particularly apt to me at present.

We have now finished collecting data for the second iteration of this business survey, and that will be published today on our website, today being the day that I recorded this rather than the day you might be watching it. We plan to run this business survey at least monthly while the economic effects of the spread of COVID-19 continued to evolve and continue to be of relevance to the community and policy makers.

We are also running a household survey, which we have dubbed Rapid Acquisition of Population Information and Data, and the acronym is RAPID. This household survey is collecting information on changes in the employment circumstances of members of the household and on the impacts that the restrictions on social gatherings are having on people. We have completed enumeration of this survey late last week having received responses from just over a thousand households. As with the business survey, our aim is to publish the results as soon as possible.

Along with these new surveys, we are identifying new sources of data to help shed light on the evolving implications of the spread of the virus. These include using near real-time scanned data from the supermarket chains to reveal the community's changing expenditure patterns. Toilet paper, anyone? Single Touch Payroll data from the ATO to shed light on employment and income patterns across industries and interactive mapping technology to provide geographical information on the distribution across Australia of older people and those with health risk factors relevant to their susceptibility to infection with the virus.

So that's another example where these data are available, but their use has become extraordinarily more important because of the virus, and that's important not just to governments, but for community groups. For anyone who is seeking to plan where resources need to go, it's important to know where the people who are most at risk are. As I said, never waste a crisis. So those are some of the things that the ABS has been doing in immediate response to the spread of the virus, and one of the things that I would leave you with is that in circumstances like this, it is sometimes possible to do things in weeks that would take months under normal circumstances.

Right across governments and in the community, there's a huge hunger for trying to help with this to do things that are actually helping with the response to this crisis. So I would encourage you all to think about innovative ways that you can use data in the organisations in which you work, and you may well be able to break down barriers that would normally take you months very quickly because if you've got a good idea that's going to create a data source that's going to be valuable, then lots of people are going to be keen to help you in this environment.

Having talked about what the ABS is doing often in collaboration with other parts of government, so the Single Touch Payroll, we can't do without the help of the Tax Office, and there are several other examples here where we've got help from others for the data we've been collecting. So having talked about that, let me turn to some of the longer term messages that I talked about in my IPAA address on the 11th of March. I just want to run through some of those because I think they're also very relevant to people who are in the Graduate Network.

So that talk for those of you who've had a chance to look at it was called The Promise of Data in Government. I sought to do a few different things in that talk, and I want to talk about each of them. The first was to spend a bit of time talking about a development, which is not completely new but is increasingly important, which is building integrated data assets, which multiply the value of data by putting together data that examines a problem from several different perspectives.

I run through in the talk a series of data assets that have been built. I mainly focus on the commonwealth government though there are huge benefits in expanding these data assets not just for the commonwealth government, but also for state governments. But in my talk, I talked about some of the data assets in the commonwealth government. I talked in particular about the business-integrated data asset BLADE and the people-integrated data asset MADIP, and gave some detail about the development of these data assets and how they were substantially helped along by the $130 million investment in the data integration partnership of Australia, which as been running for the three years and ends in June of this year.

Then, I came up with a four specific examples of things that have been made possible by integrated data assets. I think it's actually a task for all of us to come up with compelling examples of ways in which the innovative use of data has made something possible that wasn't possible before. I think all the people in the Graduate Data Network understand the power of data. But part of our role is not only doing cool things with data, but also coming up with compelling stories for other people about why this is important because ultimately, to get the resources we need to do this work, you have to convince people who don't have the same day-to-day experience of data that you have. So I'm always on the lookout for good examples of where a clever use of data has made something possible. I would encourage you to do the same and come up with ideas that are going to help in that endeavour.

I talked about four breakthroughs that I thought or at least things that have been made possible by integrated data that hadn't been made possible before, and I'll talk about them quickly because you can find the details in that speech. The first was researchers at the IBA using data from BLADE to demonstrate the effectiveness of a business investment tax break that had been introduced in the early stages of the global financial crisis. Often, it's hard to get a good handle on how effective these things are, and this integrated data asset made that possible in a way that it wasn't possible before. So that was the first. The second was a study which used pharmaceutical benefits scheme data to identify adverse events associated with medicines. That study identified five new medicines that were associated with adverse health effects that had not been picked up by other means. So that was the second example.

The third example is one that's particularly... which I think is particularly compelling, is a change in allocating public funding amongst non-government schools. Initially, that the way that was done was on the basis of the incomes of the families in the neighbourhood of the school. But using integrated data from MADIP, it became possible to actually identify the incomes of the parents of the students who are actually at the school regardless of where they live. That's a much more accurate and fair way to base a funding formula than on the neighbourhood of the school, which is nothing like as... which doesn't pinpoint in as accurate of way the relevant piece of information. The fourth example I gave was a lot of work that Treasury has been doing looking at what has caused the Australian productivity slowdown

with a particular focus on declining market, dynamism. Again, there are insights there that you couldn't get without integrated data assets.

So my talk then went on to talk about how we get integrated data assets into the hands of trusted users, and I won't go into that, into the details of that, but I won't talk about that anymore other than to talk about the fact that we now have a virtual data lab, which makes it possible, makes it much more possible to do that.

Then, I went on to talk about the critical issue of maintaining trust, the community's trust. If we're going to use sophisticated data, including integrated data assets, we have to do it in a way that maintains the community's trust. I talk a bit about the Five Safes framework, which if you don't know about the Five Safes framework, it is something that as a data analyst, you should be aware of. It's a clever approach to maintaining the safety of data in a multidimensional way, but I won't go into more detail about that.

Then finally, the speech talked about the upcoming Data Availability and Transparency Bill, which is a work that the Office of the National Data Commissioner has been progressing and told a bit of the history of this. As far as I'm concerned, that's another strand in building the data capability of the public service. It's largely about making it easier to share data in a safe way. We have an enormous number of restrictions on the sharing of data across the public service.

Some of them are for good reason, but plenty of them are not. The Data Availability and Transparency Bill is all about trying to get to a more coherent and joined-up way of sharing data. So I wanted to run through all those arguments because I think that they are some of the core things that are going to make the use of data more compelling and are going to make the use of data easier across the public service.

So let me now talk about some of the things that we are beginning to do that are new and just happening. So I've got three examples, which I'll talk about. Again, they come from the ABS, but in several of these examples, they're done jointly with other departments or agencies. I think that is very much a theme of data. It's often the case that data initiatives work better if you do them collaboratively across government.

So the first one I want to talk about is innovations in the ABS's address register. So the ABS has a register of residential addresses across Australia, which is as complete as possible. But of course, residential addresses change, and every quarter, about half a million residential addresses have a change of some kind. What do I mean? I mean, new houses get built. That's one thing that happens, but houses get demolished, or subdivided, or turned from individual houses to medium density. In order to keep the ABS address register up to date, we have to quality assure these. If you do it by hand with the benefit of aerial imagery, you can do about 34,000 addresses by hand with the resources we have available.

So in collaboration with CSIRO and Data61, the ABS has developed a machine learning version of this, which is an automated image recognition software, which can look at the aerial photography and make a much more educated guess about what happened to that block. We can now quality assure five times what we were able to quality assure before, before we had this technology. So this technology has just gone into production, and we're now sharing it with the UN so that other national statistical organisations can use it. So that's the first example.

The second one is similar in some ways, and this is teaming up with Geosciences Australia. It's to do with, again, using satellite imagery data and machine learning methods to identify specific crops. So you can imagine that once again, a machine learning algorithm can do this much more quickly than an individual. So we are working with Geosciences Australia to create an open-source national land use and land cover map that will be a key input into National Accounts' statistical release. This will eventually

assist the government in responding to floods, fires, and droughts because you've got a much better idea about specific crops. So that's the second example.

The third one that I'm going to mention is to go back to Single Touch Payroll. So let me just tell you a... which is a data source that the ATO runs, so just let me tell you a little bit about that. Single Touch Payroll is a new way of reporting tax and superannuation information to the Tax Office. Businesses report employees' payroll information, salaries, wages, pay as you go withholding in super to the ATO each time they pay employees through the Single Touch Payroll software.

From July 2018, large employers with 20 or more employees were brought into Single Touch Payroll, and small employers with 19 or less employees started on the 1st of July 2019. So the ATO is already receiving over 2 million employer records and 20 million employee records every month. So this accounts for 95% of large employers and 50% of small employers. So this provides enormous potential for the ABS to fundamentally change how we use administrative data to compile official statistics. As I said, we have got access to this in the context of COVID-19, but the longer term potential for this to be able to give us high-quality, near real-time information about what's going on in the economy is simply enormous.

Okay. Those were some of the examples. I'm really trying to whet your appetite for the opportunities that are opening up. I guess it's always exciting to work on data, but given the near universal understanding now about just how valuable data is not just in the private sector, but in government, this is a fabulous time to be starting your career in a discipline in which you're going to use data in increasingly sophisticated ways. It's even an exciting time to be Australian Statistician. Thanks very much.