



INSTITUTE OF  
PUBLIC ADMINISTRATION  
AUSTRALIA

02 6154 9800  
PO Box 4349 Kingston ACT 2604  
admin@act.ipaa.org.au  
[www.act.ipaa.org.au](http://www.act.ipaa.org.au)

ADMINISTRATION

# TRANSCRIPT OF EVENT

## GRADUATE DATA WEEK DATA IN DFAT

**Dr Jenny Gordon**  
Chief Economist  
Department of Foreign Affairs and Trade

24 March 2021

Enquiries should be directed to Caroline Walsh on 0413 139 427 or at [caroline.walsh@act.ipaa.org.au](mailto:caroline.walsh@act.ipaa.org.au)

JENNY GORDON: Thank you. It's a real pleasure to be here today to talk to the data graduates. I work with DFAT. I'm the Chief Economist there, and I head up the Office of the Chief Economist. And in my Office of Chief Economist, I've recently set up something that I'm calling the Data Analytics Section, because I think it's a signal of just how important data is to everyday good policy analysis and practice. So, DFAT is not a data agency. It never thinks of itself as a data agency, anyway, so it's not like the ABS or industry or agencies that have very large data collections, primary data collections, and that have to manage those and use them appropriately. But we are a data agency; we just haven't realised it yet, because data is what we use to inform our decisions. We use it to work out what the priorities are, and we use it to support service delivery.

JENNY GORDON: So, what I wanted to do today was take the opportunity to just tell you a little bit about how data is used within DFAT, and then some of the things that the Office of Chief Economist is doing to try and strengthen the ability of DFAT staff to use data in really effective ways. So, I'll give you an example of consular. So, we don't have a lot of administrative data collections. Administrative data collections, as you all will know, are where government collects data for particular programs and management, but then you're left with this really rich data resource that we can use for a whole bunch of analytical purposes that give us a lot more insight than we'd have otherwise. So, with consular data we have passports, so we know all the Australians who have been issued with a passport. We also try and have lists of everybody overseas, so when you travel, ideally you will register. And so you'll register on the DFAT that you're overseas and we know where you are.

JENNY GORDON: And clearly with COVID, our consular has been incredibly busy. And data has been really key, because we had a registration process for people who are overseas who wanted to return back to Australia, and that number keeps growing. So, if you start thinking about what's the distribution of arrivals into the list of people that you're going to want to try and get back to Australia, to try and understand where that demand is going to be. We need to know where those people were, because if you're organizing extraction flights, where are you going to be able to fill a plane? So, if you don't have that data, and you don't have that data well-organised, it's really hard to do that.

JENNY GORDON: And so DFAT had that data but it wasn't well-organised, and so it's been an absolute struggle, and some stellar efforts by some of our staff. And I've now pinched one of them for the Office of Chief Economist, so I'm very pleased to have him involved, but they pinched one of my really good grads, actually, to go and work in that area. So, they were able to start putting together this data in far more useful ways, and then also use it to communicate to government where we were at. So, how many people were on the list? Who are we getting home? Where were we prioritising? And then if we had had the data really well-organised, we would have been able to add additional value, so you can add additional layers, like where are the COVID risks highest? What do we know about the data about when we book a plane, that the people are likely to actually turn up and get on it?

JENNY GORDON: So, we need this sort of additional data. What is the COVID risk of those people likely to be coming in? Because we know the COVID information of the areas that they're living in. So, once you've got a really well-organised set of data, you can add additional layers of data to try and understand what is the best policy option, and the best way of managing things. So, consular data has been really critical in DFAT being able to bring Australians home. And if we had had really well-organised data, we could have worked out, with all the state caps and the like, how we could have tried to optimise the number of people we bring back, given those state caps.

- JENNY GORDON: And we could have worked better with the airlines. We've worked very effectively with the airlines to try and bring people in, but if you've got well-organised data, it just means you can do it a lot faster, chew up fewer resources of people who are trying to put data together in spreadsheets. Make less mistakes, because your data is well-organised and you're using code, rather than spreadsheets with formulas in them to try and work things out, which can get a little unstable, as you all know. And so, it's really a valuable thing that we're looking to do in the future, is they have our consular data much better organised so that we can actually add a lot more value to it, and be much more responsive by being able to just get there faster in terms of the information we need.
- JENNY GORDON: Another major area of data, which sits firmly within the Office of Chief Economist, is our trade and investment data. And so, trade and investment data is collected by the ABS. Very interesting data. We get it at a fairly disaggregated level, so some of it we then confidentialise by clumping it up a little bit so that when we make it available to people, you can't identify the particular firms involved, which is an important part of good data management. I should actually mention, just thinking about consular data, well-organised is really important, but privacy was key. So, maintaining the privacy of the people who are in that data system was actually also super critical, so we can't ever forget that. And confidentiality is another aspect of data when we're looking at data that is around businesses and the like.
- JENNY GORDON: So, what we're doing with this data is we provide generally just a baseline understanding of where we're at in trade and investment in Australia. So, for all our sort of... The ministers are very interested in this, all the policymakers are interested in where we're at with how our trade and investment is going, and we even understand what our relationships are with the rest of the world. So, if you're looking at sending a delegation to a particular country to negotiate a trade agreement your starting point is, what's our current trade with them? What's our current investment with them? If you don't have that data, you can't answer that question. So, a lot of what we do is provide that kind of baseline starting point data.
- JENNY GORDON: And so, we then start adding layers of analysis on top of that. Interesting questions we've got, like the China-US trade deal. So, you had this wonderful trade deal between the China and the US where they had these great targets of how much trade China was going to buy from the US, and we looked at it and we went, "Oh, yeah. That'd be really interesting." And so we've been monitoring it, but we're not just monitoring how much China was actually achieving the targets of the agreement that it set with the US, we're also monitoring whether that has had any impact on Australian trade. So, have we had trade diversion? Have they been buying US meat and not buying Australian meat? And then if they do that, where does that Australian beef go? Does it go into other markets that the US had previously been supplying or not? So, trying to understand and unpack can only be done by tracking through the data.
- JENNY GORDON: We also access the global trade database, which is a big database where pretty much you take the kind of ABS Australian data, but you get it for every country. Fascinating trying to match exports from Australia to a particular country who've got their import... they think they're importing from Australia. So, data is always really interesting about, what does that tell us about how they count things? What does that tell us about how we're counting things? Can we align the two and work out what the differences are? And that's a very helpful thing to do. And in fact, with economies that don't report particularly accurately, some of that data of what we know we're giving to them is really valuable, because that's how... if we know all the other countries who are giving something or they're exporting something to them, we can work out what's actually much more likely to be going on in those economies. And that tells us a lot of things.

JENNY GORDON: So, that tells us things around what are the industries that are growing in those countries? What are the industries that are contracting in those countries? So, you get a lot of information by understanding the data that you've got in place. Clearly, China's trade actions; another really interesting area in terms of trade data. What has happened to the barley that we used to export to China? In fact, it's gone elsewhere. So, then you can ask the really interesting questions of which markets did it go into, right? So, we saw this lovely sort of big fall in exports to China but a big rise in exports elsewhere. But what was then the price penalty that the suppliers, the farmers, suffered as a result of losing that Chinese market?

JENNY GORDON: So, it helps us ask the right questions and it helps us answer a lot of those questions as well. And it helps us to go and work out where we should be targeting things. Supply chain vulnerabilities. That's a really tough one to unpack because you've got to look at the links of where did we import stuff from? But where did they import the inputs into it from? And are they also dependent on one supplier somewhere? So, it's actually, you kind of bring together data but also a lot of information and knowledge about what that industry looks like in order to actually be able to say something sensible.

JENNY GORDON: The third area of data that we look at is really, supporting our development assistance program. So, Australia has a \$4 billion dollar a year program of official development assistance. The government recently expanded that considerably to support vaccine distribution in our region, in the Pacific and into Southeast Asia, but also to help with recovery. So, monitoring the ODA spend is a huge source of admin data. So, we have a program called AidWorks, and it's not just contract management but it also monitors performance. It monitors the risks, so you're collecting data around... every project reports back on a set of safeguards around child safety, around child labor, around other sorts of health and safety risks, around corruption. And that gives us a really interesting source of data to start doing some analysis.

JENNY GORDON: My area also does evaluations, and evaluations are very qualitative, often, in nature, but you can actually start using artificial intelligence and machine learning to draw out sort of common patterns in some of the evaluations that you're seeing. So, we're starting to just explore those. So, even though I said DFAT's not a data agency, I have a staff member who actually knows how to do this stuff, which is just marvelous because it opens up our opportunities to do far more sophisticated and interesting analysis. So, there's the administrative data, and again, it's getting your data well-organised so that you can use it, is a really important part of a data professional's role. But then using it is sort of the next role, so that's the data analyst. And then there are people who collect the right data, too, in the first place. And one of the things you find with AidWorks is that the people who put the data in often locally engage staff in the posts, and they don't necessarily see the end point in the use of the data.

JENNY GORDON: And so, if you can feed back the use of the data to them, they suddenly go, "Oh, it matters that I put that data in accurately." So, it helps us build the quality of the datasets we have, and we've got the checks and balances in place so that we can really understand what's going on. So, in addition to the administrative data, we use a lot of secondary data to really understand what is going on in the countries that we are providing development assistance to. And that is very, very important. So, we use a lot of World Bank data, IMF data, but also we get a lot of information from our posts of data that we then have to keep confidential because it comes from the country's government or surveys that are not publicly available.

JENNY GORDON: And so, one of the really important things for DFAT is to have really clear classifiers on whether the data is something that we can, as long as we're respecting privacy and confidentiality, get that and put it out into public domain, or whether the data is something that we actually have gotten from someone with only permission for internal use, and then we can't use that data externally. So, we have to be really careful about that. But again, if you store your data well, if your data is well-organised, you put a label on it that says... it comes up with a red warning sign saying you can't publish this particular data point. And so, one of the things that we're working on is a data management framework, so the Office of Chief Economist sort of said, and I think there's been attempts before, said DFAT's kind of data is not terribly well-organised. Right?

JENNY GORDON: And so, one of the really important things is to make sure that you have a number of... of kind of the way you think about data is as a system across the whole organisation. And you have to have principles around how you treat, think about data. So, if you think about data, you collect it. First, you design how you collect it. So, the design of how you collect it really influences the kind of data you get, the quality of data you get, it's ease of storage, it's comparability. You have to design that collection well when you're collecting data. You have to then clean it, usually, particularly if you're getting data from secondary sources. You spend a lot of time trying to work out, is it accurate? Is it not accurate? How do you check it? What do you do? How do you make sure that everything's been put in the right spot? Have you got good metadata for it, so that it actually explains what the data is.

JENNY GORDON: And then you obviously have to store it. So, you want to store it in a way that protects privacy, right? But you're going to have to also store it in a way that you can extract it for use. And then you have to ask that question; if you're combining it with other datasets... and this is where it gets really interesting in data science, is: does that reveal other things, when you combine it with other datasets, that might violate privacy or confidentiality? So, that's a really kind of an interesting question that the data scientists, like Data61 and places like that, are working on. They're trying to understand if I combine these two, can I do it in virtual space and get the results without ever actually seeing the individual datasets put together? So, if somebody knows your suburb you live in, and your age but not your gender, then there's a huge number of people that could be. But once they now know your gender, or the fact that you have a particular color hair, that starts narrowing down.

JENNY GORDON: So, when you add layers of data, often from very diverse sources, suddenly you couldn't discover that you've actually... can identify with them. So, you can bring the data together in ways that the data is never actually sits together physically. It's brought together in virtual space, the analysis is done in virtual space, and the results are published or the results are available from this, but the data never gets put together in a way that allows you identify people. So, there's lots of really interesting things we have to think about, and data principles are effectively telling you how you're going to collect data, how you need to store it, how you need to document it, how you need to make it available for use. What are the rules around that, and then what do you do with it when you're done? A lot of people forget that bit.

JENNY GORDON: If you've collected data on the guarantee that that data will not be used for any other purpose than the purpose this once-off use, or this set of uses, then you should destroy it. And you need to guarantee that that data is then wiped clean, destroyed, and no copies are made. And so, ABS is very, very active in this space about assisting organisations, not just government but organisations generally, on how do you fulfill all those issues around data. So, the first thing is principles. The first pillar of a data management framework are your set of principles around how you treat data across all those dimensions, including your analysis, so you need to make sure your analysis is used in a truly robust way. So, you don't mine your data for the answers you want; you use your data to test your hypothesis and is your theory correct. So, that's part of your data principles.

JENNY GORDON: Then there's a set of governance around data, which obviously is trying to make sure the principles get enforced, and so you have to have... is a decision-making structure and a regular review of where you're at and what's going on in data. Then you need a sort of what I call the IT infrastructure. So, where do you store data? Do you have an enterprise data warehouse? Is that how you're storing it? If you're storing it on a bunch of Excel spreadsheets, that's probably going to be go back to your principles and you decide that's not very good of... satisfy those principles. How are you going to store that data? What are the protocols around actually extracting it? What are the tools that you can use, the IT tools? So, can you use open source software to actually extract something, or is that going to cause you some security problems? There's a whole set of issues around the kind of institutional arrangement architecture, and that's also a big investment decision, too.

JENNY GORDON: So, if you're investing in a whole bunch of IT systems, you want to make sure that they serve as many purposes as possible. The public service has a rather bad reputation of having lots of bespoke bits built to do particular data needs that don't talk to each other. And that is really, really problematic. So, we would really like to... the data management framework's idea is that you enable... sometimes you truly do need bespoke data systems, particularly for administrative data to do what they need to do, but can you extract that data with protecting privacy and confidentiality and the like, to be able to combine it with other data, to be able to do really good policy analysis? So, that is the sort of issues around the governance, works out what you can do, but then have you got the technology that allows that to be supported? So, you haven't just got silos that can't talk to each other, and then silos where one falls over and nobody knows how to fix it. That's the other problem, too.

JENNY GORDON: And then this fourth pillar is capacity building. Really, really important. And that's, I guess, why you're data graduates, because we need you desperately, because young people are much more data savvy than old people. We haven't necessarily grown up. I mean, I'm an economist so yes, I used to teach econometrics. So, kind of, I'm a data nerd; let's face it, as this talk has probably demonstrated, but it's a good thing to be a data nerd, but there's a lot of sort of people who are really... they don't understand it and they get confused by it. And so, building the capacity at that end to use data, and use data to illustrate your point, to reinforce your argument that you're making, and in DFAT this is really important because, good examples are things like, "Oh, the amazing work that we were doing to pivot the development program to help with COVID-19." So, you have a lot of development activity happening, and then your country gets COVID.

JENNY GORDON: And so, what is the impact of COVID on that country? Where are the things that you need to really address? How do you understand the impact on poverty? How do you understand the impact on domestic violence, on children going into school? Where can you get the data that's going to tell you those sorts of things? And then if you're thinking about designing your programs, which are the ones that are most important? And so, you need to be starting to ask yourself the question of, if domestic violence has gone up by 10% or 10 percentage points, which is often what you'll hear, is that 10 percentage points on a domestic violence rate of 60%, which unfortunately is the case in some places, or is that a 10 percentage point on a domestic violence rate of 10%?

JENNY GORDON: They're very different situations requiring very different strategies, because in one, domestic violence is widespread, common, and you'd have a different set of strategies than you might have, if it really is at the margin where you'd be trying to identify what are the real levers that have caused that massive increase. So, you need data and information to tell you where to put your effort and where to make that investment.

JENNY GORDON: So, I think that's probably it from me. Enjoy your being a data graduate. I'm sure you will. And I hope that you get to do lots of really different, interesting things. And remember that no longer, I think, does having a data nerd label on your head mean that you're confined to just the technical end of things. I think the way you bring information to bear, the way you think through analysis, is absolutely critical because that's what we really need for good policy making in this country. Thank you.