TRANSCRIPT OF PROCEEDINGS

Maximising Value from Data: Navigating the Opportunities and Challenges

Half-Day Conference

Gandel Hall, National Gallery of Australia

17 OCTOBER 2019

Transcript 2: Session One

Dr. Jill C.:	So let me join the panel on stage, but before I do, let's give each panellist, if I could, just each of you a brief five minute opportunity to give us your point of view on our topic, which just to remind folks, is on data usage and community expectations. So I'd like to ask each panellist just to give us a five minute opportunity for your perspective about the topic, and Rosie, perhaps we could start with you?
Rosie Hicks:	Sure. Thanks very much. So the Australian Research Data Commons... I'm sitting absolutely in the university research sector, and I'm feeling slightly outnumbered this morning. And what a wonderful opportunity it is for me to reach out, and that's going to be one of the themes here, for success. It's about reaching out between these different sectors, and that includes most of the people in the room this morning with the academic sector, and also with the commercial sector as well.
Rosie Hicks:	So the Australian Research Data Commons is federally funded. It's part of the National Collaborative Research Infrastructure Strategy, which is a portfolio of research infrastructures, covering activities that span right the way from the oceans to the stars. And the Australian Research Data Commons is in a wonderful position, acknowledging what all agree this morning, I'm sure, that data is absolutely underpinning. So we influence right across that sector. And the purpose is to provide Australian researchers with competitive advantage through data.
Rosie Hicks:	Now what does that mean? Well, we've just heard very simply about how important it is to share data, and bring data from many sources together for meta analysis. There's also the huge issue with research integrity, which is hitting the headlines more and more, including this morning. So that was very topical for us. So the use of public sector data by researchers, as we've heard, is informing evidence based policy, a very virtuous cycle when we have the data flowing from government to the researchers, and back again. Medical data, personal data, key concerns we'll hear a lot more about this morning I'm sure, but we're also talking about weather data, farming data, administrative data on the use of services.
Rosie Hicks:	But going in the other direction towards commercial data products, and I'm being very careful in the distinction there between data itself, and the data products. For example, one non-medical, perhaps non-sensitive example I'm going to share with you, is the use of data for farming and drilling right down. How do we help sheep farmers forecast fly strike, and improve protection of sheep? Well, that work's been done, so of course I'm being really specific here about how data can inform in all directions, if we can break down these barriers. To do that, we do need to be moving from ad hoc, and the data legislation that's at the forefront of these conversations this morning, is certainly going to help address that.
Rosie Hicks:	We've heard that we need infrastructure, that's not just the hardware, it is the people. What I want to emphasise is it's the cultural change. And in the world of the Australian Research Data Commons, and not just the research sector in Australia, but globally, there's a framework that we're talking about, and that's the use of fair data, and perhaps this is the one word that I want to leave you

with today, fair. It means findable, accessible, interoperable, and reusable. One of the reasons this is such an important tool in our discussion, is what does accessible mean. It means as open as possible, but as closed as necessary. So being able to put that degree of control, and actually comfort to the people that we're working with, I think is extremely important.

Rosie Hicks: Now I mentioned that fair reaches beyond Australia, it is discussed globally, and I can share with you that some work's being done on estimating the cost of not having fair data. So in the EU, this has been forecast to exceed 10.2 billion Euros per year in economic loss. And then of course, we've got the social impact of improved health outcomes, we've got the environmental impact of reducing natural hazards.

Rosie Hicks: So fair data, it sounds like a wonderful thing, what are some of the barriers to that? Well, it is education, it is shared systems, it's absolutely shared language, looking at the data versus data products. We hear people talking about, "Well, we need more access to government data," what does that actually mean? So helping create the framework to drill down and answer those questions to create those standards, is critically important for us. So for the Australian Research Data Commons that's our business, looking at how we improve the research infrastructure around data for improved outcomes.

Dr. Jill C.:   Rosie, thank you, appreciate those comments. Ian, I might go to you now for a few minutes of perspective.

Dr. Ian O.:   I'm wondering how to follow that. So I'm New South Wales Government's Chief Data Scientist. The first ever Chief Data Scientist brought in to run an analytic centre charged with addressing wicked policy challenges. Wicked policy challenges are those which are complex, subtle, and ultimately have people's behaviour at their heart. I arrived at that role working through Nokia, helping to plan and optimise networks where people were absolutely going through a data driven revolution. Some time at CSRO, where I saw amazing things, absolutely amazing things done with data, including the work of the astronomers. A brief stint at a financial market data company, where people were desperate to get data, and it really was an arms race as to how much data you could get. And landed in NSW Government running a data analytics centre, which had no data, no data scientists, no compute platform, no resources of any kind, but a piece of data sharing legislation, which was being tested and would soon be put up, and a set of problems endorsed by New South Wales cabinet, that ranged from fire and rescue response times to compulsory third-party insurance, to domestic and family violence, and out-of-home care reform.

Dr. Ian O.:   And it's the most unusual environment I've ever worked in. Across that range of challenges, there are many, many people in government who care deeply about delivering better outcomes, who are also quite used to doing things in a particular way, and who deeply distrust data driven approaches, when in particular, non-traditional data sets are used. And those concerns range from data quality issues, to data availability issues, to the quality of analysis, to having sufficient context to actually understand an insight generated from data, to how far that data is away

from the real world, and how complete the picture is that's created. And I think Abigail talked about some of the things we know, we know, and some of the things that we know we don't know, trying not to stray too far into a particular analogy there.

Dr. Ian O.: And it's been an interesting journey because, what we've discovered along the way, is that we can really do some amazing things around optimising trains, around identifying fraud, helping bring down the cost of compulsory third-party insurance. But when we get to the people centred data challenges, those ones that matter most, where people care most, where the challenges are most complex, that the philosophy of bringing lots and lots of data to bear on a challenge, and taking approaches that may well work in a less sensitive, more deterministic, or more explainable environment, really don't quite work the same.

Dr. Ian O.: And over the course of the last four years, we've made really good progress building data assets around out-of-home care reform, longitudinal data sets literally for millions of people, all de-identified, containing thousands of variables, which are providing some amazing insights, which people flatly reject, which people say cannot be true, and, which every single argument about the data quality, the data use, the data capture, the data bias, the data interpretation, the context required, are put forward as reasons why that can't be the case.

Dr. Ian O.: So one of the issues around community expectation is that, we are slowly but surely realising that data doesn't provide answers, it provides insights, in particular when we're dealing with that very complex world of human services, and difficult sensitive circumstances. And helping to inform those insights, which can then be backed up by experimentation or other methods, is a way of slowly but surely winning trust with our own colleagues, as well as trust with people who actually deliver those frontline services, as well as trust with people who are impacted by those frontline services. And it's quite a patient process that we go through, when we're talking about human centric services.

Dr. Ian O.: So in terms of community expectations, we have realised that, what we can technically do, what we can legally do, the risk appetite of the organisations we work with, public perception of what we're doing, and the perception of public perception, don't line up. And we're only four years into our journey. We've got quite a long way to go, and we are having an impact inside New South Wales Government, and now genuinely in the real world, but there's quite a lot of learning that we have to do. Some large slices of humble pie that we've been eating along the way, but also some upskilling and some mutual education that we're doing with our colleagues, and now more recently with NGOs and with frontline service delivery.

Dr. Jill C.: Thanks a lot Ian, Marion, over to you.

Marion Hemphill: Great. We're all here today, because we recognise the importance of data. And whether that's in the health sector, or for government, or in the commercial sector, we're all on the same page. That the more quality data we get, and we're better at analysing it, we'll be able to make better decisions, and I do think the

community accepts that to an extent. The difficulty is you've got to get the data in the first place, and particularly in the health sector, that's personal data, and very sensitive data.

Marion Hemphill: And one of the things I wanted to share with you today, from my perspective having been through a data breach with my organisation, is the impact on trust and social licence if you have a breach, and what we've really learned about community expectations regarding the data that we collect and our use of it, and who we share it with. The biggest thing I think I'd like to share, and it might be obvious, is that there's a big gap between what legislative requirements are, and what community expectations are. Your legislative requirements are your minimum, they're your absolute base. The community's expectations are much, much higher, in terms of the limits around the use of the data and what you can do with it. They're really putting more of an ethical framework around it.

Marion Hemphill: So people come in and they consent to giving blood, or they consent to using an app, we consent every day to give our personal data to somebody, but we don't read it. We don't take on board... We might think about it in the back of our mind, but really, we're just wanting the app, the new game, or we're just wanting to get over with donating blood, or receiving medical treatment. Apple's the perfect example of this, because it takes apparently around 12 to 15 minutes for the average person to read the Apple terms, most people hover on the button for six seconds. I don't know what we're taking in, in that six seconds, you might as well just go for one.

Marion Hemphill: But it does sort of show that, consent is a bit of an illusion, and when you have a breach, the people's whose data and personal information is impacted by that, they don't care what your terms and conditions say, they're not interested, they're interested in what decisions you made with their data, and a relationship of trust. And I guess for government, the expectations around government are probably higher than for the private sector. There is issues around trust for both the private sector and the government, but there does seem to be a sort of a mistrust that data's going to be used for other spurious... Like for My Health, for example really, people should be thinking, "This could be a fantastic thing to use for getting precision medicine for me and helping with my ongoing care," but a lot of people are really suspicious that it's going to be used to remove services, change funding, or privatise some services.

Marion Hemphill: So, I guess, I can't believe that as a lawyer I'm saying in front of you, don't really worry about the law so much, it's really about putting an ethical framework in place, and deciding, not what can we get away with, with the data, what should we be doing with the data, and what will the people who are giving it to us be comfortable with, and feel is right.

Dr. Jill C.: Thanks a lot Marion, and your comment's actually picked up a point. I'd just written down prior to your speaking, about the difference between social licence and legal obligations, and you've just illustrated that point perfectly. Can I leverage off that, and we'll kick off a period of panel discussion for a few minutes, and then open it up to audience Q&A. So if you've got some questions, or

comments, or issues you'd really like perspectives from this panel on, make a note, because we're going to have time very shortly to actually collect... to have an opportunity to have that discussion.

Dr. Jill C.:   But if I can just pick up where you finished off Marion, as you've noted, certainly data is really quite a commodity now, and increasingly so, and a valuable one, and one that's got risk around it for all the reasons you've identified. And certainly, many of us work in environments, where issues such as the risk of cyber security breaches, and cyber attacks, have the potential in a data context to result in enormous breaches and disclosures, which in turn could have significant impact on the credibility of the activities of many of our organisations. Can you comment, Marion, and then other panellists I know, will have views on this too, on practical steps that any organisation should be thinking of on that protection side, that risk management side?

Dr. Jill C.:   So clearly there's a system's infrastructure piece, but there's probably a human dimension to this, and probably a process or system protocol piece to this as well. So I'll be really interested in your views on that?

Marion Hemphill: Yeah. I think the best way to prepare for a breach, because you're going to have one, is to have one, because I feel we're in a relatively stronger position than we were two years ago, or nearly three years ago now. It's really about making sure that you've got your systems in place, you've got your policies, everyone knows what their role is, everyone accepts that this is going to happen one day, and so they're ready for it, but practise. I think it's fine to have policies and things written down, but until you've been through the processes of it, you never really grasp what's required.

Marion Hemphill: So I think having simulations and crisis management exercises, utilising the policies that you've got, also get someone who is not within your area, to come in and have a look at your policies for you, a fresh pair of eyes. And then when you've gone through your exercise, have a debrief, and figure out what you would do differently, and then how you should be building them. And sort of do that regularly to make sure that you've got... You want to create a muscle memory, so that when it does happen, you can move forward. Because my experience for a breach is that, there is a very human response, and it's a response... well, it's actually not a response, it's a reaction, when you first hear about a breach, you do, you sort of have a wee panic, and you have a bit of a fight or flight response.

Marion Hemphill: You don't want to be trying to invent something in that state of mind. What you want, is a nice security blanket to hold onto, so that you've already thought through how you're going to work through these processes, so you can move from your reaction to your response relatively quickly, and you can take that confidence forward. But yeah, practise, practise, practise.

Dr. Jill C.:   Thank you. Ian or Rosie, did you have comments on this issue about protection of data, and the credibility impact of course that can occur in the event that it's not protected?

Dr. Ian O.:	Sure. So I think a breach, it's a really good idea. I went through a simulated breach recently, and it was terrifying. It was about half an hour into the simulation, when we stopped thinking it was a simulation and really had that flight or fight reaction. One of the points though, a breach is one part of a concern about data use, and some of the things we've been looking, the voiced concerns always put to us when we're asking for data to do analytics projects, the voiced concern is typically privacy. You can't have that, because even though de-identified, because there's so much of it, you might be able to re-identify these de-identified data sets, especially because we link so many together.
Dr. Ian O.:	The unvoiced concerns though, are really not about sensitivity of data subject, out-of-home care, domestic and family violence, not so much, because we've got protocols and governance around use of those, they're the unvoiced concerns around, "What are you going to do with this data? What are the implications of finding those insights? What will that mean to me? What will it say about the job we've been doing? What will be the consequences of using it? How much context do you need to accurately utilise that result? How much harm could be created, is that harm reversible? And, are the decisions explainable?" We've mapped out around 10 sensitivities related to data use, as opposed to concerns about what happens if you get a data breach. One of them is, what happens if there's a data breach.
Dr. Ian O.:	There's also ultimately the concern of, what's appropriate use of this data, and what happens when things go wrong? So I mentioned harms and reversible harms, and we've done quite a lot of work with an advisory group, who essentially take the models and systems we've put in place, and poke holes in them. So it's a really good idea to get other people to look at your systems, and what happens when things go bang problem, what happens when there's an outcome which is adverse as a consequence of that data analysis. What do you do? Can you put the genie back in the bottle, or can you reverse that harm? Or, do you acknowledge that there's residual harm, and build other mechanisms around it?
Dr. Jill C.:	Thanks, Ian. Rosie, did you want to add anything to that?
Rosie Hicks:	Well, I think the key point I would add, that actually goes back to a lot of what we were seeing in the initial poll results, is from my perspective, we're looking at how we incorporate this kind of discussion in the training side. So I'm going right the way back to early career researchers, and helping them understand implications, best practises, and being able to answer in terms of their own research agenda, some of the issues that Ian's raising.
Dr. Jill C.:	Thanks, Rosie. I might just build on your comment too about capability actually, just in terms of another question, which might really useful to get some panel perspectives on. All of you I know, and many in the audience, will be often thinking about the future of work. And so, that broad theme about how work is changing go forwards, what occupations are becoming, or likely to become less common, or less required, what in turn might be more in demand. And so many will say that professions around data analysis of course, digital skills perhaps more

	broadly, but certainly data analysts, statisticians, econometricians, people who can do machine learning and things like that are going to be increasingly in demand in time.
Dr. Jill C.:	Do you want to comment a little bit on what that means for capability, in terms of how we build it in data analytics? So, is this a concern that we have in Australia, that we don't have enough data analytics capability that we're building coming through? Is it something that we need to invest in more generally in our working population to lift everyone's level of data comfort, perhaps a little bit more than it is at the moment? So, I'm really interested in your comments with that future of work lens about data capability in the broad data analytical capability, and that understanding of data given that it's sort of a new and emerging growing commodity.
Rosie Hicks:	May I?
Dr. Jill C.:	Please, yeah.
Rosie Hicks:	May I start there? So there are two points that I'd like to share in this. We've had a lot of discussion in our sector about this absolute issue, what do we need more data scientists, and I'd like to put a different perspective. We need people that are bilingual, we need people that are able to work with the data, but then also with the discipline as well. I'm talking about people that can work with the astronomers and the data, can work with the social science community and the data, and we're not just operating in an isolated environment.
Rosie Hicks:	So there's been some discussion about creating a new stream, a new field of research in data science, and that's actually been rejected, because we don't want... rejected in some conversations. I'm sure that they remain passionate advocates, but I would reject it, because I don't think we want to deepen that split between where we're applying the data, and the understanding of how to use it. So that would be a really key point.
Rosie Hicks:	And the other one I think we struggle with, is there's an absolute unanimous recognition that we need more skills, we need more data, we need more of this magical substance that's going to cure everything. Hang on a second, could we be a little bit more specific about who needs what, and particularly, whose responsible for providing it? It's an enormous challenge, and unless we work out the different roles and responsibilities across the sector, we'll keep looking for the magic wand that will address it. So I'd like us to increase the sophistication of discussion around digital skills.
Dr. Jill C.:	Thanks, Rosie. Ian, or Marion?
Dr. Ian O.:	So I somewhat agree with Rosie, somewhat disagree, but that's okay.
Rosie Hicks:	It's our job.
Dr. Ian O.:	That's right. So-

Rosie Hicks:	We're boring otherwise.
Dr. Ian O.:	There are some pretty impressive tools coming onto the market, and have been around on the market, which do automated machine learning. And what that effectively does, is make a data scientist more productive, or it allows you to do much more exploration. A data analyst could more of what was upon a time considered to be data science, and they will continue to get better over time. I think a general awareness of what data can do, how it can be interpreted, how it should be used, a little bit like electrical safety, is a skillset that we should be introducing right at the very beginning of the education system, and of course, those who want to specialise can. But just an understanding of what is possible, what's not possible, what would have to be magic in order to actually work, I think, is an important skillset.
Dr. Ian O.:	Ultimately, there's a blurring between the world of computer science, and the world of people who do data stuff, and so I think broadly speaking, there's a stem sort of challenge, but it's not unique to this space. But the ability to understand and translate... that bilingual translation between the world of data and the world of outcomes, or the non-digital world, which I think is still called the real world, that translation is still going to be very, very important, and increasingly so if we're relying on data.
Dr. Jill C.:	Thanks, Ian. Did you want to add anything, Marion?
Marion Hemphill:	Yeah, I really liked Rosie's use of the bilingual view, and focusing on the languages, and I think, we constantly, in the human race, have to move forward. We've gone from blacksmiths to mechanics, and now we're moving into artificial intelligence, and it's not just for specific sectors to be focusing on data specialists, and the uses of AI, it's going to be impacting us right across the board. So I think it's up to every industry to really do a lot of work on research and development, and horizon scanning, and having a look at what's next, because you won't be able to prepare an initial period of time.
Marion Hemphill:	We're going to have to look off... make a prediction as to what's going to happen in 10, 20 years, so that we're educating and training the right people with the right skills now. And I hope that we look more globally for that, and harness what's going on in different economies and jurisdictions around the world, and we're not just sort of repeating and using different languages to go in different directions.
Dr. Jill C.:	Thanks, Marion. So there's already been, I think, some really good points from that discussion, let's have an opportunity I think, to take questions from any folk here who might have something that they'd like to put to the panel.
Cameron:	Good morning all, my name is Cameron, I'm the Media Advisor at the Office of the National Data Commissioner. I'm fairly new to all this, so now I'm looking at data as a tool that you can actually talk about. So it's a whole new elevation to me. One of the things that really struck me this morning, Dr. Oppermann, you spoke about, with the data you're doing, you're coming out with some great

answers, much more detailed new insights, but you're running into this cognitive dissonance, have you come back and say, "Actually, this is what the data's telling us, of people going, "Oh, no, that can't be true, that doesn't... I've already situated the appreciation and you haven't come back with what I was expecting."

Cameron: Could you just talk, and perhaps the panel members, on overcoming this cultural thing of people not wanting to accept the information that comes out when it is at odds with what they'd already planned the outcome to be?

Dr. Ian O.:   So, thank you for the question. Don't mind if I just lie down on the couch now and do a little bit of therapy. We have made every mistake you could imagine about trying to deliver results to people who don't want to hear them, or who don't believe them, from the approach. I'm sure we're all in government. Giving results directly to cabinet before the relative secretary has been briefed, bad mistake in terms of building relationships. Telling people that we're going to report to cabinet, and giving them time to come up with counter answers, or counter justifications, handing results through a slide door, and saying, "Okay, they're yours now, and off you go," and then disappearing into a slide. Or, when we moved into treasury, talking to the treasurer about it, and the treasurer going hunting for the cost savings that we apparently found directly from the agency that we were working with.

Dr. Ian O.:   So all of those things are the wrong thing to do, in case you were wondering. The best way that we've been able to approach this, is to warm people up slowly, and to, when we identify a result, help people go through the change process. Because ultimately, it's a change process. Sudden change in dramatic circumstances, which matter, are very, very... often very badly responded to. So helping people warm up to it, and doing that in way that we not only show results, we don't ever say why a result is... sorry, not say why result has eventuated. We never say why, we say is, and we say is, with a level of confidence.

Dr. Ian O.:   So we can say, "This is our result, and we are 85%, 95% confident of the result," and then we'll test it together, and we'll test it from a context perspective, and we'll test it from a, "Have we done all the right things from an analytical perspective? Is there possibly a problem with our data?" We'll go through the whole lifecycle of data, and if the result stands, then we might test it from a different perspective. But if that result stands, then everyone's got more confidence that it really is the case. Now that's slow, that's really, really slow.

Dr. Ian O.:   And we got started, as I mentioned, we had basically no resources, and we were keen to deliver results, so we were really pushing out results that we were pretty confident about, but without that contextualization of warming up, and it works really well if you're looking at a train timetable. It doesn't work well when you deliver a result, which talks about rates of sexual abuse for children under 12 in out-of-home care, it's an alarming result. And so, now we would take the time from the very outset, when looking at a project like that, to actually ensure we've built in place those mechanisms to patiently build trust in the result, and the insight, with the client.

Marion Hemphill:	I've got a slightly different view on that, and I think that it comes down to the fact that data tells, stories sell. And I think that when we're trying to persuade someone to make a change based on evidence that's produced by data, the evidence alone isn't going to do it. There has to be the human element to it, because you're not selling data to data, you're selling data to a human. And so I think that... Going back to the last question about making sure we've got the right specialisations, I think there's a real role for marketing and communications around the data piece, because it doesn't actually sell itself.
Dr. Ian O.:	I completely agree, completely agree.
Dr. Jill C.:	Would you want to add anything to that Rosie? You're fine? Any other questions? There's a couple. We might just go to the lady over here, and then I'll come to the gentleman in the middle.
Cathy:	Hi, my name's Cathy Vosloo. I'm from the Department of Health, and also a PhD student at the Crawford School of Public Policy, and we're looking at how you take big link data and turn it into policy and social impact. I'm just starting, so I've got a long way to go. So my question is, in the kind of analytic process, the knowledge creation process with data, I'm just kind of riffing off what's already been said, how many people need to be involved in that process? What kind of skills need to be in the room? What kind of views need to be represented, do you think, in that process?
Dr. Jill C.:	Any takers?
Dr. Ian O.:	Well again, we started with nothing, and so anything more than that is a good start. We also started looking at the least sensitive problems, and the first ever problem was fire and rescue response times, and that was a great example of actually winning over a commissioner through storytelling. So my apologies for not raising that point. A storyteller, someone who can translate, is a really, really important skill if you ever want any analytical result to go anywhere.
Dr. Ian O.:	We brought in... As our projects became more serious, and more sensitive, and more important, data governance was actually the most important skillset. We were doing... We needed to demonstrate we were doing the right thing with the data, we were applying it in the right chain of custody, in the right chain of governance to the data sets. And then analytical capability came after that, because if you don't seal the holes in your boat to the extent you can, you don't get ferry passengers around. Terrible analogy, but anyway the point is, that bit has to be right.
Dr. Ian O.:	Then we built... We started adding more and more people who would help address one of the challenges of getting data. People don't share data, unwilling, unable, not allowed. The unable is, it's expensive, it's difficult. We got offered data on magnetic tape at one stage, we said, "We'll get back to you on that." But the cost, and effort, and expense of getting data and feeding the analytical engine, was really an important aspect. We used a lot of PhD students, by the way, for that. And then we kept building around that loop.

Dr. Ian O.:	So our very first team consisted of master's students actually, and a data engineer, a platform's person, and a governance person, and we just built around that loop.
Dr. Jill C.:	Do you want to comment?
Rosie Hicks:	I think I'll just add one comment that hasn't come up so far in the discussion today, but is really key, particularly in the health space in Australia, that increases the number of people required and the complexity of the problem, and it's the cross-jurisdictional limits that you see particularly in that space. And perhaps contrasted with the astronomers that have come up multiple times this morning, and make a point there, that it's about the maturity of data dependence in that particular field, where we're seeing the astronomers out in front, and the humanities arts and social sciences at a stage where we can do lots of good work by transferring between disciplines.
Rosie Hicks:	I think there's a huge opportunity in the next few years, in bringing up that baseline.
Dr. Jill C.:	Do you want to add anything, Marion? There's a colleague in the middle here, I'll just get a microphone to you.
Joe Walsh:	Hi, so I'm Joe Walsh from the ACT Data Analytic Centre. I kind of wanted to pick up on a bit of a theme that's coming out, which is money. Obviously, half of the people here want more resources to do good work with the data. I wanted to get the panel's thoughts on, I'm sure everyone here has examples of that time where how good use of data saved a lot of money, paid for itself many times over, how do we move the conversation from how data is expensive, to not having data as expensive?
Marion Hemphill:	Great question. I'd start with, it's not what it costs, it's what it buys, or what it saves. Like in the medical space, with My Health for example, my understanding is that a lot of tests that are done in Australia on individuals, like blood tests, pathology tests, scans and that sort of thing, about 30% of them are repeats, and are unnecessary, and that's because data isn't being shared. So that's not great for the individual, because it's their time, and also they're having to have a needle stuck in their arm, or whatever the test is, but it's a waste of an enormous amount of resources that could be used for something else.
Marion Hemphill:	So I think it's actually, this might be a circular, but finding the data on the savings and using that to, it's not about the cost, it's about the benefit, and making sure that your money is well spent.
Dr. Ian O.:	So I'll join two themes together. So when the data analytics sent us data, every project result had to deliver an economic value. And we delivered some economic value estimates, which were pretty small in many cases. When we finally bid for budget, the minister responsible gave up his speaking spot in cabinet, so that I could tell cabinet about what we were doing. And getting back to Marion's point about storytelling, I waited outside the room, it was 7:00 AM, I could hear the cabinet having a good old chuckle inside, and thought, "This is good, this is a

Dr. Ian O.:   And 30 minutes later, I was still telling stories about the things we discovered in the data, and the impact on real people, as well as some of the economic benefits. And to be fair, the economic benefit was of interest, but that is not what people responded to. It's not what the members of cabinet responded to. They responded to the very human stories about what we'd found in the data. What happened from that, unfortunately, one minister got embarrassed, so back to my earlier point, don't do that, because his team hadn't briefed him, and he was sitting right in front of me when I was asked to tell the story about what we'd found in his portfolio.

Dr. Ian O.:   But what did happen after that is, all the ministers wanted to have a one on one conversation about what was going on, and I was speaking to then treasurer, and now premier of New South Wales, and saying, "What we found, we think it's sort of 100 millionish here, and maybe another 100 million there." And her response was, "So nothing serious then." And we'd just asked for our funding bid, and I said, "Well, have I got a deal for you." We moved into treasury, we now work... we have been working on some billion dollar programmes, and now we've moved to this new customer service cluster.

Dr. Ian O.:   But all the way along, we have been looking for economic benefit as a second order priority, noting that without that we don't get to fly, but that's not what wins the hearts and minds of people who genuinely care.

Speaker 8:   One question I wanted to put to the panel is, how do you distinguish good data from bad data? There is an assumption that the insight derived from data is always good, but as we heard from Professor Payne, if the data that supports the analysis doesn't reflect the population, it could be a misleading insight. So how do we distinguish good data from bad data, and whether bad data can be fixed by AI and machine learning techniques.

Dr. Jill C.:   Thanks, Yumin. Rosie, did you want to kick off?

Rosie Hicks: I'll start by pausing to consider what does good data mean. And I think this is an extremely important question around data quality, and that one of the key points we need to consider in whether data is in fact or not, is what's the provenance of that data. And I think, rather than saying it's good as a subjective measure, it's of known provenance, and therefore we have a degree of confidence, whether it's high degree or low degree, in the decisions that we're making using that data. So the data provenance, and ensuring we are working within frameworks that don't see a data collection in isolation, but a data collection with particular attributes supporting it, and enabling the quality of the decisions to be made at known degree, is what we should be targeting, rather than the subjective good or bad in that particular data set.

Rosie Hicks: And I'm going to give an example to ground that for us. So if we look at some of the data that's available today, the complexity of the census used to capture

environmental data, for instance, and we look at what we had 100 years ago for capturing temperature, or wind, or soil quality perhaps, the quality of that data, whether its good data or bad data, it's the data that we have to make a longitudinal study. So understanding more about the provenance, how it was sampled, what degrees of tolerance we had on those measurements, is what we have to work with, and incorporate into our models.

Dr. Jill C.: Ian?

Dr. Ian O.: I'm going to suggest first of all, great loss to ABS that you've just recently retired. I think that's a very, very profound question, and I think I would ask the question slightly differently, and not use the term good or bad, because they are quite subjective. We believe philosophically, that all data is useful in one way, shape or form, and all data is a way of seeing the world. All data is biased, all data is incomplete, all data does not have full coverage. And acknowledging that, the question is then, what can you do with it, or what could you rely on that data for?

Dr. Ian O.: There's an example of a project we did around builder insolvency, and we looked at three different sorts of builders. And there was a builder field, which was actually quite predictive of whether a builder would become insolvent, and that was their size. And when it was missing, that was actually a strong indicator, that this particular builder was in trouble. And not because it was missing, the fact that no-one was really paying attention to what was happening to that builder, may well have been the reason, accept, we don't say why, we just say is. Anyway, that data was still useful, and so we effectively created some metadata around it, but we couldn't rely on it the same we could where we had a more accurate higher quality data set, which informs.

Dr. Jill C.: Thanks, Ian. Do you have anything to add?

Marion Hemphill: I think it's a key question, that I really wish I had an answer, but I think it's a question that we have to keep asking, and testing the data that we have, and getting assurance and understanding its limitations.

Allen Watford: Hi, my name's Allen Watford, I'm from the Department of Human Services. We've spoken a bit today about, I guess, combining data, managing what data we can get, I was just wondering what the panel sees as the biggest barriers to making sure that we use this data ethically, especially when it comes to trust from the country and the public? Thanks.

Marion Hemphill: I don't know if it's a barrier, but the biggest problem is good intentions. They're really focusing on what they want to achieve with data, by using it... You've collected it for one purpose, but you're wanting to use it for another. But that's a really great purpose, and why wouldn't someone expect that and consent to that? But with these great intentions, we try sometimes to put ourselves in the minds of the person whose consenting, and what they might expect, and what they would be comfortable with, and I think we look at it through a very rose coloured lens, and are probably a bit optimistic.

Marion Hemphill:	So I think a barrier is perhaps our singularity, and looking at the data we've got in front of us, and looking through our own lens. And I think a way to break through that, is to get assistance with ethics in the privacy and the data space, and don't make your own decision. Put a proposition forward, but then ask your ethics committee, and if you don't have one, ask yourself why you don't have one. Put it to someone who is able to stand back and look at the situation more broadly. But yeah, the lack of consent is a big barrier in itself.
Dr. Jill C.:	Ian, did you want to comment?
Dr. Ian O.:	So, as a general statement, one of the big achievements that New South Wales has made, is talking about outcomes, and being very transparent about what those outcomes are. I was on a panel once upon a time talking about the safes model, which I'm sure everyone's familiar with, safe project, safe people, safe data, safe setting, safe output. And the response back from the audience was, safe for whom? And it was an aboriginal woman, and she said, "None of the stuff you're talking about, seems particularly safe to me."
Dr. Ian O.:	Being able to describe what we were trying to achieve in terms of outcomes, being very transparent about those outcomes, and then measuring those outcomes in an increasingly sophisticated way, and making that measurement transparent, is what's helping with this trust building exercise we've been carrying out with our own partner agencies inside New South Wales Government. And that outcomes thinking, is becoming more prevalent, because we can have the debates about whether these are the outcomes we should be seeking, we can have the big pea and the small pea politics argument about them, and we can be transparent about what we're actually measuring as far as those outcomes are concerned.
Dr. Ian O.:	The really and exciting stuff that's starting to happen right at the moment is, through the collaborations with the Australian Digital and Data Council, are all States territories and Common Wealth are coming together and talking about outcomes they want to achieve, and building cross-jurisdictional data assets. And particular, the first really significant one, is a national NDIS Data Asset, built along the lines of what we're doing with out-of-home care, described in terms of outcomes, rather than inputs. And obviously, that will be a very sensitive, very important data set to help underpin reform of the NDIS.
Dr. Ian O.:	So we're getting there, but part of the answer is that transparency, and part of that transparency is talking about outcomes.
Dr. Jill C.:	Rosie, did you want to add anything?
Rosie Hicks:	I'll simply add that, the idea, or not the idea, the use of ethics committees in the university environment of course, is well established. And I say with a bit of a smile, perhaps we need to make sure that those ethics committees have some bilingual people. And by that I mean, people that are deeply engaged, and across the types of discussions we need to be having with the data, and as we move into

Dr. Jill C.:	Thank you. And I think we've got one last question right at the very back. Yeah, go for it.
Dr. Helena Z.:	My name is Dr. Helena [Zabortsifan] from Department of Industry Innovation and Science. I'd like to thank panel for addressing vital data issues that shape public policy, and life of people in a way, well, lives of people. My question is prompted by several insightful questions asked by Professor Abigail Payne, and while we've addressed the issue on data limitations, and data omissions, I'd like to turn it into perspective, data flexibility.
Dr. Helena Z.:	Professor Payne referred to an example, what if an individual shouts the sky is green, well, we all know sky is blue, and some people might believe it is green. I thought that during rainbow, some beads of the sky actually are green, and maybe some people who have different vision issues do see sky green. And sometimes we just don't take into account some other perspectives on data, and if we fix policy using specific data, then it is very difficult to correct it for those individuals who are not covered by the data. So how important it is to still leave space for human corrections, and not to be over dependent upon specific views on data?
Dr. Jill C.:	That's a great question, and as a native Queenslander, and with our enormous hail and electrical storms I can assure you, the sky gets very green in Brisbane on a regular occasion. But I'll be interested in panellists who might have a view on that.
Dr. Ian O.:	So, a great question, and as a native Queenslander, I've also seen the sky being green. I think it depends what you're using the data for, what you're using the data driven insights for, and whether or not those data driven insights are actually going to be used to drive a decision or an action and something happens as a consequence. And that really speaks to the sensitivities around what are the consequences of that action being taken.
Dr. Ian O.:	We've got automatic trains coming to Sydney, which is fantastic, driverless trains, and if you're an automatic door, on an automatic train and you sense that there are more than an expected number of people, so the door stays open for another second, that's pretty low context, pretty low consequence. Closing it a second earlier, might have consequences, but keeping the door open for another second, doesn't really have significant consequences, versus if you are looking at a drug trial, or if you're looking at some service intervention for juvenile justice, contact with juvenile justice for example, there's no way that you would take an insight driven by data, and automatically implement an action based on that. It would be absolutely insane.
Dr. Ian O.:	So that degree of context required, or human in the loop required, again, split out by consequences, or harm, or explainability, or sensitivities... inherent sensitivities, I think is where we really need to ensure, that if someone's saying the sky is green, we go in there and make sure that we've got at least a different set of

perspectives on whether or not the sky actually is green, and whether it's green for some people and not for other people.

Dr. Jill C.: Did you want to add anything to that, [crosstalk 00:51:14]-

Marion Hemphill: Yeah, I think it's an interesting question, that if someone comes up with something, and it's going back to what you were saying about an unexpected result, and I think even though we can't just rely on data 100%, we do have to still use our base understanding. So if someone says the sky is green, and we don't think it is, we have to sort of have that courage to speak up and say, "Why do you think it's green?" And then sort of go back and, I guess, not be afraid to challenge the outcomes on the basis of all the data that we've collected over our lives, and I guess, our own intuition, for want of a better word.

Dr. Jill C.: Do you want to add anything, Rosie?

Rosie Hicks: Perhaps we could consider that before we had the data available on the colour of the sky, we were relying on the anecdotal evidence, and then we would be distilling it back to the same question, who we're listening to? So whilst it's very fraught, difficult and careful conversation we need to have, we can distil it back to, what did we have before data, and are we able to improve our decision making as a result of new information, as well as the multiple sources that are available to us.

Dr. Jill C.: Great question. And I'd like to say, an enormous thanks to our three panellists this morning. It's been a fantastic discussion, and I'm sure you'll agree with me on that. We would actually like to present each of you with a small gift, which are... Thank you very much, Sarah, which are award-winning handmade chocolates from our local chocolatier, Jasper and Myrtle here in Canberra. While I do that, could you join with me in thanking the panel.