Opening remarks to the opening plenary of the 2020 Data for Policy Conference
London
15 September 2020
Sir Peter Gluckman, Koi Tū: Centre for Informed Futures.

I want to focus on the issue of trust.

Using data is nothing new for governments – whether in the form of the doomsday book or in establishing their tax intake. Of course, we are now dealing with the age of increasing use of big data and algorithmic policy decisions, and so the discussion takes on different dimensions. Different societies have different levels of trust in their governments' use of data and information. In no small part, this is because of the varying nature of the political State, democratic or authoritarian, and the perception of where power lies: the thought that information is power is real in the minds of citizens.

It is clear that data can help governments make better decisions, but there are also dual-use considerations. Data and the associated infrastructure, worryingly, can also be used as a means of control, so these discussions have real implications for citizens. Think of the objections in many countries to even having an identity card.

The concept of privacy is changing in the big data world we live in. Social media encourages us to share material we would not have had two decades ago, and privacy itself is a concept for which there is large cultural variation, and it is far from the only issue.

Surveys and focus groups in many countries, including my own, and a recent study from Imperial College show people are less trusting of a government having their data than a company. This surprises many people, but it is a very consistent finding and suggests a deep issue over social license for government use of data and perceptions of who can have access to their private data. People understand at some level, perhaps naively, the nature of their bargain with Amazon or Google – we allow these companies access to our data which we know will be monetised through micro-targeted ads, in return for a service. But when personal data are collected by a government, what will it be used for? Certainly, there is the tax and benefits system, but in general people do not link their own receipt of services to the giving of their data to the State. What instead are the concerns of those who worry about this? Indiscretions revealed? Loss of autonomy? Rarely do they really believe that giving a government their data will genuinely be returned in the form of better services. In some countries, even filling in a census form is fraught with non-compliance and resistance.

NZ saw this when it developed the Integrated Data Infrastructure, which links information about every citizen across broad sectors of government and was designed to help governments make better decisions about investments in the full range of social services.

But it was implemented without adequate social license, which together with misstatements as to its purpose and the almost inevitable bureaucratic error, jeopardised public trust. As a result, it was unfortunately politicised, which further undermined trust, limited its scope and its value to policy makers. From being best in class, the real potential of the system to ultimately dissect out what works and what does not work across the social services has not yet met its full potential. Yet given the limitations on any country's budget using citizen level data to understand what services work and in what context, be it in education, health, social housing, welfare, justice and so forth, could be of enormous value to society. Doing so, however, requires genuine partnership and trust between state and citizen. It seems to me that this should not be a partisan political matter; indeed, robust interpreted data creates space for more honest values-based debate. There is an opportunity we may yet squander to advance economic and social sustainability, but the policy community is yet to fully understand the issues that must be first addressed.

It is reasonable to assume that the implications of this type of situation have now flowed on to Covid-19 responses, as the concerns and trust deficits have played a role in inhibiting the introduction of supplementary digital contact tracing in many countries. Other issues have emerged in these experiments in the use of data in social policy, such as the debate over data sovereignty for indigenous people in New Zealand. I think this is a proxy debate for deeper issues of disempowerment, labels, discrimination and fear of misuse of data in ways that would reinforce those biases. What is clearly needed is trusted and independent oversight over government use of data. The issues are about more than privacy, and more than traditional ethics. Data can do a lot to improve the human condition, but governments have been reluctant to understand that their own use of data needs to be subject to trusted and principled oversight, which in turn requires an exercise in co-development with citizens.

We must remember to ask the question: what is data? Data is not knowledge; data is the aggregate of what we can measure with all the flaws about its collection as we try to study phenomena and gain a better sense of our reality. That is, we *turn* data *into* knowledge by organising it and finding some meaning to it. Increasingly, data are put into models of various forms. Those models try to describe systems. But those systems always have some unknowns. And the models we use are, inevitably, designed on the assumptions and path-dependent structures that are built into how those systems might be conceptually modelled to describe reality. But the assumptions along the way may not be obvious. The interdependencies can be non-linear and exist in ways that can only be guessed. For one thing, the data can be non-representative. We tend to try and model open systems as closed systems. All this cries out for the need for data to be married with expert interpretation and analysis before it is called evidence, let alone knowledge. And the policy and political community need to appreciate these issues.

Too often, uninterpreted data are turned into dogmatic statements of certainty. We must remember that numbers and graphs can be remarkably rhetorical. Rhetoric is important, but we must be careful to understand the power of big data and models to be very

rhetorical yet not good for communicating uncertainty, for integrating knowledge beyond the data, and for dealing with hidden biases.

Covid-19 highlights these issues, with epidemiological models being extensively used. Graphs have become a major form of communication. But the quality and purpose of models and presentation have been very variable – some are trying to force decisions, some are attempts to understand behavioural elements, others recognise diversity in behaviours, and yet others are very simplistic and based on naïve assumptions about human behaviour. Despite this, little has been communicated about uncertainty. Often, details of the model and its assumptions are not available and peer review is not the norm until well after the model is used, if at all. Uncertainties beyond the model – and, indeed, beyond the data – often remain largely unstated. In part, this is because factors of importance may be left out. Data sets may be biased, as we have seen in the case of facial recognition.

Remarkably precise claims are being made for predictions by some from these models, which are then taken up in the counterfactual with great dogmatism by politicians. Of course, models have been critically useful, but less hubris and more reflection is needed.

We are early in this journey with big data, and as a society we appear both enthused by the hype associated with it, but also rightly concerned. Let us think through the limits of it, and be more demanding of understanding the issues any big data claim brings in the social sector and what kind of linkage to expert interpretation and oversight is needed.