

NEW ZEALAND DATA FUTURES FORUM

NZ IS A WORLD LEADER IN THE TRUSTED USE OF SHARED DATA TO DELIVER A PROSPEROUS, INCLUSIVE SOCIETY.



NEW ZEALAND'S DATA FUTURE

FULL DISCUSSION PAPER



NZ DATA FUTURES FORUM

#NZDFF



Executive summary

The New Zealand Data Futures Forum thinks New Zealand needs to have an important conversation. The world is changing, connecting us all like never before. We are sharing and using data about ourselves, our communities and activities, our work, our devices, and our environment – often without realising it. Does the “big data” revolution herald an age of possibility or a “big brother” future? New Zealand needs to adapt – and how we adapt matters. It will influence our future economic and social prosperity.

Benefits and opportunities

The new data environment gives us a richer and deeper understanding of the world, generating a range of benefits and opportunities for New Zealand:

- **Competitive advantage** through innovation and a world-leading data environment
- **Business opportunities** through new markets and an expanded knowledge sector
- **Better public services** arising from data-driven efficiencies and better targeting
- **Better places to live, work and play** – smart cities and optimal use of natural resources
- **Everyday life transformed** through automation, personalised services, and informed choices
- **More open transparent government** generating greater trust and empowered citizens.

Challenges, tensions, and risks

This new environment also exacerbates concerns about access to, and control of, data and the potential for competing interests and misuse. Risks of misuse include:

- **invasion of privacy** through misuse or mishandling sensitive personal details
- **invasive use** where individuals are targeted; can be merely annoying or actually harmful

- **discrimination and exclusion** from services based on correct or incorrect information
- **malicious use** for criminal purposes, including identity theft and fraud
- **big brother**, where the line between legitimate state power and individual liberty is crossed.

Next steps

New Zealanders have some difficult choices to make. The opportunities have the potential to bring huge benefit to New Zealand, yet the risks are very real. The NZ Data Futures Forum thinks the only realistic choice is to find a well-managed way through the middle – to try and maximise the benefits and minimise the harm so that we can adapt to the future in a safe and effective manner. That's why the Forum exists.

We want to know what you think:

- What kinds of benefits and opportunities should we be aiming for?
- How can we make it easy to share and use data when it is required?
- What risks and challenges need to be managed?
- Are there privacy-friendly ways to access and use data?
- How can we protect the interests of New Zealand and New Zealanders?
- What will a sustainable and adaptive environment for data sharing and use look like?

Visit our website www.nzdatafutures.org.nz for further information.

1. Introduction

The world is changing in an important way

In the last few decades, digital technology has become a ubiquitous part of people's daily lives. It has already led to major changes to the way we live, work, and play. Many people no longer use paper and pen to send a letter, but instead send an e-mail, Skype, or chat online. They also share photos, create online networks of friends and colleagues, watch movies online, participate in virtual games, and store their information in the cloud. Financial transactions are made directly from home or via smartphones using Internet banking. The person behind the checkout counter at the supermarket no longer needs to manually record the price of a product, but merely scans the barcode. This barcode not only records the price automatically, but also tells the shop owner what the product is and stores this information alongside all of the other purchases that buyer makes that day. The supermarket knows exactly what was in the buyer's shopping basket. If the buyer is a member of a loyalty scheme, then what the buyer purchased is linked to the buyer's identity and linked to other transactions the buyer made throughout the day, such as at the chemist, with the bank, at the petrol station, or the furniture store.

These examples illustrate lives that are becoming increasingly "digitised". It is going to increase. As well as use of the Internet, people will leave new kinds of digital 'breadcrumbs' behind. When people swipe their card to get on the bus, this will inform the bus company about the use of bus transport to help the company with capacity planning. When people drive under a toll road and a photo is taken of their licence plate, this information can be used for automatic payment or billing for use of the road. That same information is used by the Transport Agency to predict future road use. The smartphones in people's pockets increasingly have a number of "senses"; the latest models have an accelerometer, gyroscope, and GPS, and can also measure temperature, humidity, air pressure, light intensity, and magnetic flux. They can sense people's movement. Together, this sensing of movement, speed, and location can reliably determine whether people are walking, catching a bus, or driving a car. Smartphone sensors allow a whole new market in exercise and fitness "apps" to help people plan and monitor their exercise.

Medical researchers in England are collaborating with people who leave their smartphones on at night next to them in bed. The researchers are able to 'crowd source' information about dreams: a smartphone application senses when these individuals are dreaming by their movement and wakes them up to get them to write down their dream. This information is

collated from people all over the world to help researchers understand sleep and dreaming.

People are starting to wear other kinds of sensors. Ones that can track your heart rate are used to help you with your exercise. Indeed, medical professionals are investigating ways to measure blood chemistry and a variety of other information to help them help you stay alive and healthy – letting people know the optimum time to take medication.

But personal information is only a small part of the digital ‘breadcrumbs’ that can be used to understand people and the world they live in. Remote sensors can inform us about places, buildings, the environment, cars, the weather, clothes, luggage, and many other forms of non-personal information. There is a big movement to use sensing in cities; monitoring water use, traffic, use of public utilities, air pollution, the movement of buildings during earthquakes.

The cars that people drive nowadays record their driving activity, such as the steering wheel, speed, and direction of travel. This information is stored in a black box and can be retrieved in the event of an accident to help determine the cause of it.

Experts suggest there will be an “Internet of things”. Sensors will be put in appliances, in homes, and in the things we use. Our cellphones are merely the first step. It should soon be possible to wirelessly link up data from fridges, TVs, and to monitor a household’s use of power. There will be data indicating that the lights are on, whether there are people in the house, and which room they are in. Entrepreneurs will use this to provide new kinds of services, such as automation of people’s domestic routines, alerting people when the milk has gone off, automating their shopping list, and remotely ordering replacement groceries – think of the barcode reader. People are working on billboards that will use face recognition technology in order to detect a person’s proximity, look up their social networking data, and provide a relevant advert to them as they walk down the street. The same face recognition technology is already used by Internet companies that wish to now connect people to their friends by recognising both faces in the one photo. Great if you are wanting to network with your friends, perhaps not so great if you are having an affair. Face recognition technology is also used for other purposes, such as the identification of security threats at airports and the monitoring of hooligans at soccer stadiums in Europe.

“Big data” is both a method *and* the driver of significant social change

When you ask somebody from the industry, “What is Big Data?” they will usually reply that this describes the challenge that companies that collect and analyse the high volumes of Internet data face. This “big data” technically refers to the specialised tools required to store and analyse it, eg Hadoop and Map Reduce. However, this response says very little about the *significance* of today’s digital revolution and why it is the source of such excitement, trepidation, and debate.

What is true now is that more of our lives and activities are being stored digitally. What we also assume is that more of this will happen. So when we refer to “big data” in this discussion paper, we are really talking about the social changes that these different kinds of data are heralding. We are interested here in the use and misuse of all this new data when it gets used to generate knowledge.

A key aspect of this ubiquitous collecting of all sorts of data is that people wish to link it all together. When data from different sources is added together, it creates deeper understanding. In this paper, we are not just talking about disparate sources of data; we are specifically talking about linking it together.

And if all of this sounds scary then you are in good company. Like any technology, knowledge can be used for social good or to make things worse for people. Is this the start of an Orwellian “big brother” future or the herald for an exciting new age of possibility? There are views on both sides of this continuum.

There will be many benefits to increased information sharing

This new technology is already enabling us to learn faster, better manage many kinds of risk, connect up, and form communities of shared interests.

If researchers can use private sector data (what people eat), link this to sensors you wear (to monitor blood chemistry, temperature, and heart rate), and link this to government health data (the genome and the medical history of people and their families), they may find patterns that can potentially predict heart attacks in advance, or find early signs of diabetes or cancer, or track and reduce the spread of Influenza or bird flu.

The Ministry of Social Development is using integrated data from various sources to better learn which of its services get better outcomes. The results mean that less public expenditure is wasted on services that don't work, and more is invested in what does work. As a result, the use of integrated data in government is already creating better outcomes for citizens, and a leaner, more effective, customer-focused service response.

Wireless sensors put into concrete slabs in new buildings will be shared with engineers (or, it would seem, available to everyone with a smartphone) who can check if earthquakes have caused structural damage – is the building safe to re-enter?

There are also benefits to businesses who want to innovate, jump on board with this new industry, and compete internationally in the new knowledge wave.

The collaborative sharing of data will improve access to information and facilitate greater participation in democracy. For example, data about fluoridation or immunisation supports informed debate and participation in the policy process; making data about the business interests of our elected representatives informs our voting choices. Good data on road transport can help communities debate new roading options.

We introduce and explore these kinds of benefits using more examples in section two.

And there are also considerable risks and challenges

However, nefarious or incompetent individuals, businesses, insurance companies, and rogue governments can also seek to use these shared datasets in ways that do not benefit the people who provided this information.

Looking ahead, we can't be certain governments will always use data benignly. The dawn raids of the 1960s, where the Government of the day was busting down doors to find overstayers, raised real controversy in New Zealand. How much easier would that task have been if they knew exactly where everybody was? It's not just governments that are a potential risk in that regard either. What if a powerful person with access to your personal information used it against you?

Imagine a public official or powerful business interest who wanted to silence a dissenting voice. How easy this might be if they could know very private and personal details of that person. Or even if a government, or business, never uses this capability, it may still have a chilling effect on people's willingness to speak out if people assume that a powerful person with access to their personal information can or will use it at some future date.

It is not only governments and businesses that may want to misuse information. What about a violent ex-partner who can track down a family in hiding, or criminals and foreign governments who want to spy or steal money and digital identities.

Nor should we forget the merely annoying. For those people who do not want highly targeted niche advertising every time they log on to read their email or walk down the street – what will their options be? If people dislike flyers in their letterbox – what about their digital cousins?

What happens to people if they decide to opt out and stay off the grid? Is this a realistic option anymore? Will people be marginalised because insurance companies, banks, and governments won't provide services unless you do consent to share your private information? Is there going to truly be genuine consent in such a case? Digital monopolies will wield considerable power. There is likely to be a power imbalance if this kind of new capability of "knowing" is not well-handled by society. This turns into a class imbalance, where vulnerable people who do not have the wealth, political support, and/or understanding to protect their digital interests.

Having and making available deeply personal information raises some troubling questions; ones that we want you to consider. Any solution will need to balance the opportunities that shared data has to improve lives with some minefield of potential misuse of that same data. We need to find some kind of acceptable balance.

We are talking about data sharing (in its many forms)

As these examples illustrate, we are talking about data sharing, which we define as providing access to existing data or linking it to other data. We are talking about many different kinds of data sharing: how citizens share data with each other, the government, and with the private sector; how the government and the private sector share data; how agencies across government share data; and how businesses share data between themselves. Data sharing can be done on a large scale with big businesses and government. It can also be about what you consent to when you share your private information confidentially with your doctor.

We are sometimes talking about the sharing of *personal* data and data related to human activities – that is, data that identifies an individual. For instance, there are sometimes greater benefits from sharing personal data, such as in situations where the service provider (eg the

GP, bank representative, insurance provider) needs to know who you are in order to provide you with a personalised service.

There is also the possibility of using data that has been “de-identified”, that is, formerly identifiable, personal data that is no longer identifying a certain individual – sometimes called *non-personalised* or *Pseudonymous data*. In technical terms, de-identification of data can happen by stripping out personal identifiers from the dataset. The use of de-identified data can still provide benefits where you don't need to know who a particular person is – eg to undertake general research. It can provide benefits *and* reduce risks to privacy at the same time. Of course, there are question about how “de-identified” data can really be, but more on that later.

We also need to consider the value of sharing data about things such as infrastructure, the weather, lake levels, cars, cellphone towers, commercial goods, buildings, or locations.

We will provide many examples of the kinds of data sharing that goes on or could go on, and also many of the advantages and disadvantages of the collaborative sharing of data. We will share some of the proposed benefits of increased data sharing. We will also outline some of the risks.

Who are the Data Futures Forum members and what are we trying to do?

The Data Futures Forum members are people who collect and use data or think about how it is being used or will be used. We work in government, the private sector, and academia. We do not reflect the whole of society. But we are citizens and have children and families who are growing up and live and work here in New Zealand. When asked to undertake this work we, as a group, unanimously decided that we should do it as an open forum. We believe that a good outcome from our work will be a realistic dialogue and gathering input from a wide range of stakeholders and champions for various issues related to big data. We think the journey of thinking about how New Zealand wishes to adapt is as important, perhaps more so, than the conclusions we draw.

Like any new technology, the revolution that we are currently witnessing can be used to benefit people but can also cause harm. It is also clear that there are a wide range of views about the proposed benefits and risks. Views vary from extreme ends of the continuum. Some of you will think that this is all good. Others, that it is all bad.

What the Data Futures Forum wishes to question is whether or not we can realise the benefits of collaborative and increased use of shared data AND protect individual, commercial, and collective interests in minimising risks and improving privacy. To what extent can we have our cake and eat it too?

We think the status quo is a poor solution. New Zealand is not able to realise the full benefits of the use of shared data. There is an upside to be had through an increased ability to share data. Nor do we assume that the current ways of handling personal and commercial data to protect privacy really caters for this new kind of future and the kinds of questions it raises.

It neither manages the risks well nor does it provide for some opportunities that will lead to significantly better outcomes for New Zealanders. At the very least we need to kick its tyres to see if it's capable of meeting New Zealand's needs in the digital future.

The move to a digital future that all New Zealanders are living through is more of a revolution than an evolution. It touches all aspects of personal lives including our engagement with society at large, with businesses, government, and with each other. This is not a technical problem. It is a question about how we want to live – given that this new technology is already changing things and will continue to do so. How does New Zealand want to adapt to a future that is rushing upon us now?

The Industrial Revolution must have seemed as overwhelming to people in the 1800s, when factories emerged, automation, cars, telephones, and railroads all arrived on the scene. It was incredibly disruptive; causing pollution, transforming the way people worked, and where they lived. There were winners and losers as old-style manufacturing, where a small family made shoes for the village, were put out of business. Societies adapted. Some adapted early and well, others late and poorly. Sometimes people had to learn the hard way that some things shiny and new weren't so good – think about asbestos. On the other side, who would want to go back to the days pre-electricity, light, and heat, and where the journey of a few miles might take a few days?

There is no reason to think that the changes we are witnessing today will be any less disruptive. In fact, these changes we are witnessing are happening faster – in years, not decades. Education is already changing; the recording, newspaper, and publishing industries all have to adapt. Whole new industries are being invented. Citizens and businesses can collaborate and network globally in ways unimaginable merely 20 years ago. Will New Zealand adapt well, reap the rewards of early adoption, and manage the real risks? Or will we follow along and hope that we somehow make it? The stakes are large for our economy and the kinds of lives we live. We are a small country but we have the right to determine the kind of digital future we want.

Insofar as this can be done, we in the Data Futures Forum think we should *actively* adapt.

But this is no “Forum” without you

Before New Zealand looks for possible solutions that are right for New Zealanders, we need to get to a better, shared understanding about what is at stake. This paper is a first step at supporting a meaningful dialogue. The goal with this paper is to outline the problem that needs to be solved:

What kinds of benefits and risks must we consider as a country, and what must any possible solution need to address to maximise the benefits and minimise the harms of increased data sharing?

This paper provides an initial outline based on our own experience and expertise. The benefits and challenges are both presented, as well as some emerging challenging questions. The aim here is to provide a good starting point for a broad and collective discussion about the possible data future for our country.

What we want from you

At this point we want two things back from you:

- We are seeking your input to see if we have canvassed the issues well enough – does this discussion paper speak to your excitement and/or fears?
- If you think we have laid out a reasonable description of the playing field, we would also like to hear your ideas about what kinds of principles and solutions might be adopted that can both mitigate the risks *and* enable the benefits of greater use of shared data.

More information about how you can participate and contact us is provided in the last section.

Case study: Sandy Pentland – The power of Big Data



“With Big Data, we can now begin to actually look at the details of social interaction and how those play out, and are no longer limited to averages, like market indices or election results. This is an astounding change. The ability to see the details of the market, of political revolutions, and to be able to predict and control them is definitely a case of Promethean fire – it could be used for good or for ill, and so Big Data brings us to interesting times. We’re going to end up reinventing what it means to have a human society.”

ALEX ‘SANDY’ PENTLAND is a professor of sociology and a pioneer in big data, computational social science, and use of technology for developing countries. He was named by *Forbes* magazine as one of the world’s seven most influential data scientists.

I believe the power of Big Data is that it is information about people’s behaviour instead of information about their beliefs. It’s about the behaviour of customers, employees, and prospects for your new business. It’s not about the things you post on Facebook and it’s not about your searches on Google, which is what most people think about, and it’s not data from internal company processes and RFIDs. This sort of Big Data comes from things like location data from your cellphone or credit card. It’s the little data breadcrumbs that you leave behind as you move around in the world.

What those breadcrumbs tell is the story of your life. It tells what you’ve chosen to do. That’s very different than what you put on Facebook. What you put on Facebook is what you would like to tell people, edited according to the standards of the day. Who you actually are is determined by where you spend time and which things you buy. Big data is increasingly about real behaviour and by analysing this sort of data, scientists can tell an enormous amount about you. They can tell whether you are the sort of person who will pay back loans. They can tell you if you’re likely to get diabetes.

They can do this because the sort of person you are is largely determined by your social context. So if I can see some of your behaviours, I can infer the rest, just by comparing you to the people in your crowd. You can tell all sorts of things about a person even though it’s not explicitly in the data. People are so enmeshed in the surrounding social fabric that it determines the sorts of things they think are normal, and what behaviours they will learn from each other.

“The promise is for financial systems that don’t melt down, governments that don’t get mired in inaction, health systems that actually work.”

As a consequence, analysis of Big Data is increasingly about finding connections; connections with the people around you and connections between people's behaviour and outcomes. You can see this in all sorts of places. For instance, one type of Big Data and connection analysis concerns financial data. Not just the Flash Crash or the Great Recession, but also all the other sorts of bubbles that occur. These are systems of people, communications, and decisions that go badly awry. Big Data shows us the connections that cause these events. Big data gives us the possibility of understanding how these systems of people and machines work, and whether they're stable.

Connections between people are really important – this is key. Researchers have mostly been trying to understand things like financial bubbles using what is called Complexity Science or Web Science. But these older ways of thinking about Big Data leave humans out of the equation. What actually matters is how the people are connected together by the machines and how, as a whole, they create a financial market, a government, a company, and other social structures.

All sorts of specialties are beginning to understand that it is the connections between people that are actually the core problem in making transportation systems work well, in making energy grids work efficiently, and in making financial systems stable. Markets are not just about rules or algorithms; they're about people and algorithms together.

Understanding these human-machine systems is what's going to make our future social systems stable and safe. We are getting beyond complexity, data science, and web science, because we are including people as a key part of these systems. That's the promise of Big Data, to really understand the systems that make our technological society. As you begin to understand them, then you can build systems that are better. The promise is for financial systems that don't melt down, governments that don't get mired in inaction, health systems that actually work, and so on, and so forth."

www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data

2. Who benefits from shared data, and how?

How could increased data sharing transform people's daily life?

One of the advantages of data sharing is better coordination of activity in people's daily life. Consider buying a house where you have consented to sharing your personal information between a range of government and private sector service providers...

Scenario: Helping you to reduce complexity and support more informed decisions

Wednesday afternoon, you're standing at a real-estate auction when your smartphone beeps to life. A locally configured i-beacon has activated an app on your phone that presents all the information relating to the property you are standing in and asks if you intend to bid. You click yes and your app preloads with maximum price and mortgage thresholds based on your financial health and personal circumstances. You are ready to go. The auction ensues and you win. At the press of a biometric button, you consent to your personal data and the auction data being shared for a few minutes with all mortgage providers who you have invited to tender for your business. The tenders are back and your smartphone makes a recommendation, which you override in favour of your preferred brand. Seconds later payments are made, contracts are exchanged and you now own a new home. Not a single piece of paper changed hands. All the data necessary to complete the transaction has been shared with your consent. All mortgage providers who were unsuccessful have their data access revoked. Better still, your insurance; broadband; home phone; electricity; gas; waste collection; council taxes; and registered addresses for tax, cars, school, and Internet shopping deliveries are all automatically scheduled for updating in time for when you move in. No effort required. And if better options are available these are presented to you, too. Even your smartphone mapping and calendar services are automatically primed for the move

– best route to work, public transport options, and travel times are all optimised and updated for your convenience. What is amazing about this is that all the data necessary to enable the above scenario to occur already exists. Today, it's locked away, largely inaccessible in vaults that are owned by different companies, organisations, and government departments. In a future where you control your data and consent to access to it for services you want, those silos and vaults begin to disappear. Imagine the possibilities.

Clearly, for this kind of example to work, it relies on high trust and accountability, consent, well-managed data sharing, and principles around user control. These are aspects that the Data Futures Forum wishes to explore shortly. But the kind of future imagined here is already emerging. In this section, we look at what kinds of benefits are possible and what is potentially around the corner – assuming that we can solve some of the challenges and manage the risks (raised in section three).

Data is already being used every day by individuals, businesses, researchers, and government to make better decisions. After all, knowledge is central to humans and human activity. For this reason alone, the potential benefits of increased data sharing are likely to be in all parts of our lives.

There will be benefits for driving innovation and economic growth, for providing better commercial and public services, for the environment, and for democratic participation and engagement. In this section, we can really only scratch the surface. A few examples are provided below that illustrate a wide range of benefits that New Zealanders and others are already pursuing.

The use of increased sharing of data across government agencies

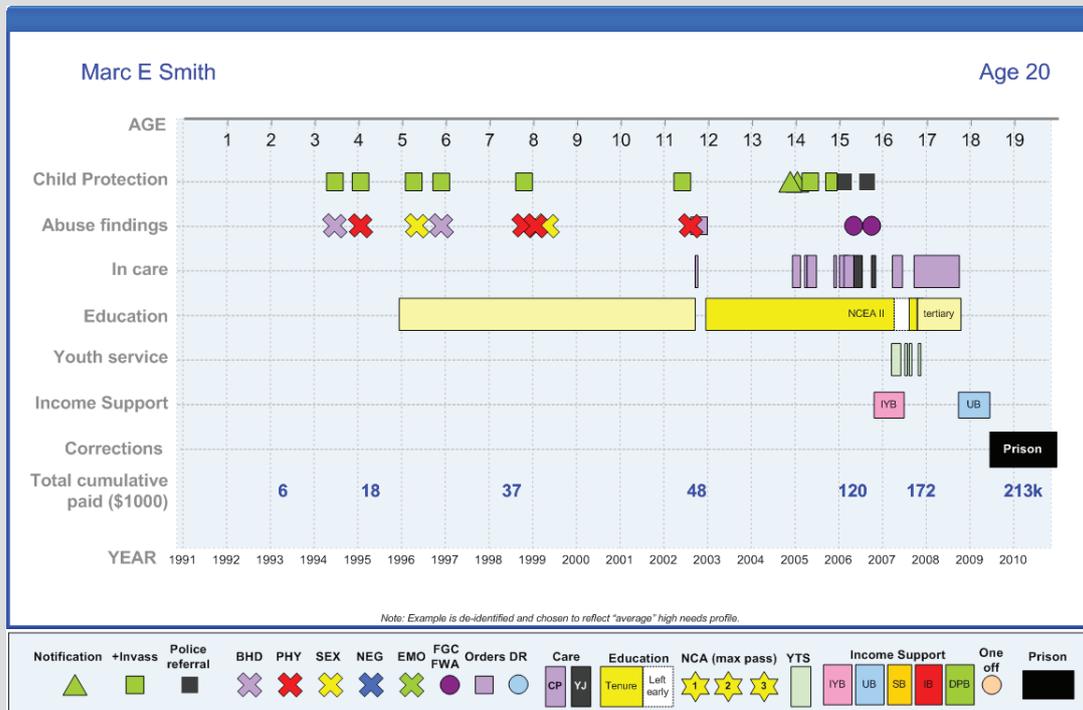
“Marc Smith” below is an example of using shared data from across government to better understand outcomes.

Example: Marc Smith

This infographic illustrates the picture for some New Zealanders when you join up the data from Child Protection, Youth Offending, Corrections, Education, and Welfare Benefits Payments data.

Marc has progressed from behavioural difficulties at age three-and-a-half, through to substantiated cases of physical and sexual abuse by age 11, on to youth offending as a teenager, eventually being taken into state care from 14 years of age. He had multiple placements with foster carers and spent time in a youth justice residence. He didn't

do too badly in education, all things considered, getting one subject at NCEA level 2. However, as a young adult he went from tertiary study onto a benefit and has spent time in prison by age 20.



The state sector has spent well over \$200,000 on Marc, yet his outcome is poor. The outcomes for the community, for his/her victims, and for the tax payer are also poor.

The profile Marc illustrates is not uncommon. Thousands of New Zealanders have a similar profile or a profile that is even more socially costly to themselves and their communities. Fiscally, a small number of people cost the country billions. This does not have to happen.

Shared government servicing data provides a crucial tool for policy and service providers to make these pathways visible, and to learn how to invest more effectively to make a difference. All of New Zealand benefits in the following ways when the state sector can have a whole-of-person perspective.

Better targeting and more efficient government services

With shared data, we can better track and make service providers accountable for outcomes. Looking at the example of Marc Smith, it is clear that outcomes usually happen after you have received a service. So if I have taken a child into state care due to child abuse, how well have I done at caring for that child? Like any parent, the outcome you want to see is that this young person is achieving well at school and is engaged in education. He or she is not trying to commit suicide or involved in offending, does not have mental health issues, and goes

on to gain employment. In this example, if we can share data from across the state sector, and all this data can be managed and used in a de-identified form, government service providers can be held to account for the outcomes of their service. For example, Child, Youth and Family can be accountable for the educational attainment of the children in their care – or better still, the Ministry of Education *and* Child, Youth and Family can have joint accountability for those kinds of outcomes. This kind of outcome focus is not really possible without shared data.

We need to research more successful ways to invest limited public expenditure to achieve better outcomes for less. Marc Smith also illustrates that government service providers often spend a lot of money late in a person's life journey, when health conditions have become chronic, a person has been offending for many years, or family violence has escalated to a point where multiple government agencies, such as the Police; Child, Youth and Family; and health service providers need to become involved. Shared government service data makes it possible to estimate the potential to obtain better outcomes earlier and can therefore inform bolder, more preventative policies.

Using de-identified shared data to have more open and transparent government

Improved transparency drives improved accountability. If everyone has access to case-level, de-identified government data, then everyone gets to know what is going on directly, through direct access to the same data that policies are made on and performance is measured on. If this were the case, then all New Zealanders would be better able to hold government to account. That is, instead of getting an annual report each year from central government for instance, you, your whānau, your non-government organisation, or your local council could look at the data directly and make up your own mind about what it says.

To some extent, this is already happening in our country. Government agencies are already working towards making more of their data available for open access and re-use under the Declaration for Open and Transparent Government. The Charities Register, topographic information, and traffic volume data are some examples of government data that is being used by people, businesses, and other government agencies for economic, social, and democratic benefits.

But that is mostly aggregate data – data that has already been analysed by somebody else, and may not answer the specific question you have. To answer your own questions, you may need access to case-level, de-identified data. There are additional benefits to open government and transparency to be gained if case-level, de-identified government data could be shared and used, also.

More diverse voices drive better policy. Greater access to existing, de-identified data and the linking of data to other datasets in a privacy-friendly way is going to create more contestable voices, alternative views, and interpretations of the evidence used in policy making. Businesses, NGOs, iwi, researchers, and private individuals can drive their own policy ideas with access to the same evidence that bureaucrats enjoy. This creates contestable policy so that it isn't just the mandarins in Wellington, or the local council, who get to have

a look at the evidence and think about what is really going on. Māori, Pacific peoples, NZ Europeans, Asians, NGOs, communities, businesses, and lobby groups might also seek to achieve similar access to evidence through shared de-identified data. This will also promote greater equity for marginalised groups who are struggling to get their message heard. Where a small or marginalised group can use evidence to put their issues on the agenda, they have a better opportunity to influence national or local government policy and to mobilise their own community-led responses to their issues. As the water example shows, we are starting to see the benefit of good data when it comes to managing water.

Greater trust drives bolder and more realistic policy. And it goes both ways, too. Where citizens and communities have direct access to the data, they can have greater confidence that government agencies are telling the truth. It will be harder to hide bad policy when people can look directly at what is going on for themselves. This is an advantage to government agencies where bolder policies can be based on deeper understanding and trust about what is really going on. All parties have access to the same data. This would be a worry only for those who wish to sweep issues under the carpet.

Example: The flow-on effect – Richer data supports policy and governance

Many Canterbury farmers are automating reporting of their water use, creating richer data that supports collaborative and informed water management.

Meters are required under the Resource Management (Measurement and Reporting of Water Takes) Regulations 2010. Metering is designed to help manage water holdings for the long term, and to assess compliance with individual resource consents that permit the taking of large amounts of water.

Under the Canterbury Water Management Strategy (CWMS), water is collaboratively managed with the tangata whenua. Environment Canterbury, Ngāi Tahu, local councils, community representatives, and other stakeholders work together to balance economic and environmental goals. The goals of the CWMS include supporting more efficient water use and involving people in decision-making. Having richer data supports these goals. Using the data automatically collected at the meters, all partners are able to assess how much water is being used by consent holders, and the longer-term effects on rivers and ground water, and they can collaboratively establish benchmarks for good water use. The data will inform future policy decisions, so water can be used more efficiently and where it is most needed, whether that is by other farmers or back into waterways that are stressed.

This collaboration is consistent with the approach taken at the Land and Water Forum, where diverse groups contributed to policy formation. The Forum concluded that good open data is extremely important to the water debate, as it supports councils, scientists, and other stakeholders to engage and comment on policy and practice.

The data provided by the meters is being used by Canterbury farmers to inform

sustainable and efficient production practices. Some farmers are combining the data from their water meter with data that comes from their irrigation system, and using the information to determine the most efficient and effective speed and spread of water. The rate of water application for each paddock and crop can be set from the office, and irrigators set to turn off automatically to avoid irrigating unproductive land or land already at capacity, reducing wastage.

www.cwms.org.nz

www.ecan.govt.nz/advice/your-water/water-metering/Pages/case-studies.aspx

Government (and private sector) use of social media

Businesses have been using social media to better understand their customers and target services. Could, and should, government-service providers do the same? Checking social media to see what people think of your government Call Centre and to learn what to do to improve it might be better for everybody concerned. Government use of social media will “close the feedback loop”, so that government agencies can learn directly from the public what is and what is not working. Inspired agencies should probably introduce this more actively, go out, ask questions, and start their own forums and blogs to harness their customers’ views to make better policy and provide better services.

Use of government data to drive innovation, industry, and economic growth

Example: Private sector use of public sector data?

One of us (who works in government) got approached by a private hospital with a request for de-identified health data from their local DHB. This private health provider thought they could do a better job of investing in rehabilitation services than the local DHB. The trouble is that they needed to analyse the data to see if this business model would work. Unfortunately, we were unable to provide private companies with anonymous data. If this had been possible, they may well have developed a more effective policy than the DHB around rehabilitation and been able to promote this idea. Who knows? Even the act of them being able to do this might put pressure on government agencies to perform better.

There are profits and a competitive advantage to businesses that have access to public and de-identified government data. After all, who says that government officials or politicians have

the last word on good ideas? Entrepreneurs of all stripes, be it these large multinationals, local businesses, or social entrepreneurs, stand to gain from better access to government data. Many of these benefits can be had by using de-identified data within tight controls around use and accountability and, in so doing, not threaten privacy. This opens up opportunities for business and social entrepreneurs:

- **to find new markets** – as the example on private sector innovation illustrates, undertaking research to identify business opportunities will be much easier with access to shared government and private sector data
- **to bid for government business on an equal basis** – businesses, iwi, NGOs will be out to make stronger cases to bid for government work on the same basis that the government bids for increased funding to provide services – eg with knowledge of demand and needs through historical servicing patterns
- **to build knowledge services** – there will be new kinds of business opportunities that are essentially involving knowledge services. If businesses are allowed access to government data, they can analyse this for their customers, eg government, iwi, other businesses, communities, lobby groups

Consider the examples below. The private sector is using geospatial data to create efficiencies for business and government.

Example: Data-driven business efficiency

Critchlow, a New Zealand geospatial services consultancy, supports organisations to realise economic benefits from spatial analysis – optimising sales and distribution networks; smart routing fleets to save time and money; identifying fraudulent activities; and improving overall organisational performance and competitiveness.

Critchlow's work includes:

Work with the Auckland Council to enable the analysis assessment, and monitoring and reporting of the Auckland region's capacity for growth. The project involved developing a model to identify the residual vacant capacity for each and every land unit across Auckland, using an algorithmic approach. It not only provided a more thorough understanding of development potential, but contributed to better evidence-based planning, and economic innovations through resource savings and timeliness.

Creating efficiencies and improving customer retention for Green Fingers, a garden-waste recycling firm in the upper North Island. Critchlow created a mapping system that better allocated new customers to existing truck collection routes. In doing so, they eliminated two hours of manual labour per day per truck; there was a 75 percent reduction in customer complaints; and a 15 percent increase in collection productivity. Through the smart use of geospatial data, economic and environmental benefits were realised.

Example: Shared open data improves urban commuting.

Researchers at IBM, using movement data collected from millions of cellphone users in Ivory Coast in West Africa, have developed a new model for optimising an urban transportation system. The IBM model prescribed changes in bus routes around the around Abidjan, the nation's largest city. These changes – based on people's movements as discerned from cellphone records – could, in theory, slash travel times by 10 percent.

The IBM work was done as part of a research challenge dubbed Data for Development, in which the telecom giant Orange released 2.5 billion call records from five million cellphone users in Ivory Coast. The records were gathered between December 2011 and April 2012. The data release is the largest of its kind ever done. The records were cleaned to prevent anyone identifying the users, but they still include useful information about these users' movements.

www.technologyreview.com/news/514211/african-bus-routes-redrawn-using-cell-phone-data/

Example: Boston Potholes App

Like every urban metropolis, Boston has plenty of potholes which are often difficult to locate and repair. So the city's Office of New Urban Mechanics came up with Street Bump; a free iPhone app that motorists can use while driving around town. Street Bump uses the phone's accelerometer to detect bumps in the road. When a car hits a pothole, the app sends information about the bump, including its location, to a database. Repair crews can go and repair the road before the hole gets too big and more costly.

www.informationweek.com/software/information-management/smartphones-big-data-help-fix-bostons-potholes/d/d-id/1105486

When you put some of this together, you start to get a sense of what Roger Denis, a social entrepreneur in New Zealand, is championing for Christchurch. New Zealand doesn't have to just follow other countries. We are leading the way in some areas.

Case Study: Sensing cities, the environment and Christchurch



There is a challenge facing the world, but it's not one that many people recognise. Today, more people live in cities than in any other environment on the planet, and this is not a trend that's likely to reverse. At the same time, however, there's a need to understand how to make cities more efficient, more livable, and more sustainable, and use fewer scarce, natural resources.

One of the key approaches to tackling this is through the use of data to peel back some of the layers of complexity that surround this puzzle. However, despite a lot of expensive marketing by technology companies, there's actually not a single smart city in the world that has the volume or coverage of data that is required. That's because it's fiendishly difficult to retrofit smart city technology into existing cities. In contrast, if you're building a city from scratch, it becomes a lot easier to do this.

Sensing City is an initiative in Christchurch to incorporate a layer of sensors, data, and analytics into the rebuild of the city. The premise is to use sensors to measure as many variables as possible such as traffic flow, footfall, cellphone traffic, noise, luminosity, temperature, energy use, and water consumption. It will also bring together datasets about the city from a range of sources that are currently held separately.

The aim is to leverage the NZ\$40 billion that will be spent on rebuilding the city to uncover greater efficiencies in the urban environment, enable citizens to participate in discussions on what's important to them, and create a new industry for New Zealand that positions Christchurch as a globally recognised hub. There's nowhere else in the world where the heart of a living city is being rebuilt from scratch, and this is the opportunity that Sensing City is seizing.

The initiative has three broad aims:

- Attract smart talent to New Zealand
- Create unique, high value, exportable, intellectual property (linked to sensor development, and data analytics)
- Attract international, inward investment from technology companies that are seeking a living lab environment

To achieve these aims, all data from Sensing City will be as open as possible in order to create an innovative ecosystem for technology companies. There will be licensing fees associated with high volumes of data usage, but for small startups and the public, access will be very cheap. An open data framework also provides the transparency that is necessary to allay concerns about privacy.

The initiative has attracted global attention, with media coverage in the Financial Times, Scientific American, and numerous mentions in construction and building magazines. Sensing City has also captured the interest of technology giants, with a range of companies actively discussing how they can participate in the project.

It's not just companies and regulatory authorities that will participate in Sensing City, the aim is also to include the people that use the city. Technology has advanced to the point that school children can build their own sensors and publish the data online. Sensing City will embrace this emerging trend as citizens start to measure what's important to them.

At the highest level, the aim of Sensing City is simple – if the country is going to spend \$40 billion in one location in a relatively short space of time, how can this be spent in such a way that it ensures that our grandchildren are more likely to stay in the country? The development of Christchurch as a world leader in an emerging, hot new industry that leverages sensors, big data, and advanced analytics is one of the few options open to New Zealand.

Roger Dennis, 2014

**SEN
SING
CITY**

Information: www.sensingcity.org

Updates: [www.fb.com/TheSensingCity](https://www.facebook.com/TheSensingCity)

Sensing City in four minutes: vimeo.com/75365337

Environmental benefits of shared data

As Roger Denis suggests, decisions based on better evidence lead to less waste. Where utilities companies, other businesses, and the government can share data, this can be used to drive efficiencies and reduce wastage. Better ability to share data is likely to drive more efficient power consumption where businesses and utilities can innovate to share data and use smart meters and other smart appliances to understand and make power consumption smarter. For example, people can reduce energy wastage by having appliances such as their heater predicting when to turn itself on or off based on people's behaviour, perhaps when waking up or coming home. Because utilities companies will know who is using infrastructure and for what purpose, it will provide the basis for better planning and management of utilities.

Promoting environmental transparency and reasonability can be done. By remote sensing water quality and making this transparent and available to anybody, people can know if it is safe to swim in a river. Indeed, if sensors are placed at different points in a river or the storm water system, this will help us to identify polluters such as dirty dairying or people flushing toxic chemicals into storm water or rivers.

Allowing, with consent, people to link your identity and government, business, and personal information means new kinds of services

What is possible if you consent to controlled sharing of your personal identity across a wide range of government and non-government organisations?

There is a new kind of service that will emerge that relies on safe use of shared personal information. This was illustrated with the example of buying a house at the start of this section, and it would require the consent of individuals where their personal information is involved. Using shared data across utilities, government, and the private sector, a service provider could automate many mundane transactions. This is great for the consumer of that service. It is also a new kind of industry and revenue stream, creating new jobs in the process.

New Zealand is investing in online authentication technology called "RealMe" that might make such ideas possible. (See: www.realme.govt.nz) There are benefits when you deal with government agencies if you do not have to find three forms of ID, get photocopies signed in front of a JP, and go and find a park in town every time you want to access a new government service. Or that if you move house you don't have to tell three different government agencies.

If people could use a similar system to manage consent to share their bank data with their supermarket data and their bills payment information, then many businesses could provide automated budgeting and a savings service. At the moment, it is frustrating for people that these services do not "talk to each other" to save the hassle of manually processing what they buy to their bank account. This sort of service really only works if people are willing to join up their data.

If this same data was then aggregated so that your data and hundreds of thousands of other people's data was also available, then this would allow you to learn from other people's

smart financial moves and mistakes. For example, you could be receiving free advice on the forecast savings rate effect if you have your particular spending profile *and* sign up for this particular Hire Purchase deal. It would probably be possible to forecast in real time your eventual retirement savings based on your own data and comparing that pattern with where other people have ended up. "Sorted" is a great New Zealand website for financial planning and retirement savings. Wouldn't it be great if it was automatic? In short, shared data will allow businesses to provide consumers with the kind of tools, resources and advice that was once only really available to people with a lot of money who could hire accountants and business advisers.

If New Zealand can "export" its brand as a safe place to share data – where there are tight controls on misuse, but it can be shared – then perhaps we can be a safe place for other countries to bring data and export the ability to build safe solutions for this kind of business. In this sense, it would be the kind of place like Geneva, Switzerland, where people come to transact sensitive business in a safe environment. Big Californian tech companies may start looking for countries that allow data sharing, protect digital privacy, and have a high trust environment where people feel safe doing so.

Benefits to researchers

Researchers in New Zealand will benefit hugely from a situation in which more datasets are publicly available to them. Some early examples demonstrate that the study of large-scale sets of newly available digital social data, created via social media, smartphones, or the use of apps, for instance, can create a wealth of new knowledge in many different areas.

For example, researchers from the UK explored real-time streams of de-identified user data generated on Twitter during the 2009 swine flu pandemic. They were able to detect an upcoming spike in an epidemic a few weeks before the official health surveillance systems in the UK and the USA did.

SAS's "Global Pulse" research used social media to look for precursors to the Global Financial Crisis and found these in social media months before the crisis hit – providing the potential to predict such events in the future.

Perhaps a more prosaic example might be the development of a scientific model based on Wikipedia page-view data and edit data, which has proven to be able to predict the success of movies at the box office.

But this isn't just about what everyone else is doing overseas.

Shaun Hendy's work (example below) currently uses open data to understand and score New Zealand cities on innovation potential and has already created interest overseas. It would become internationally world-leading if he also had access to government data (maybe tax records) and private sector data (phone network data) to better understand the complex business of innovation and economic growth for a whole country.

Case study: Professor Shaun Hendy on data-driven policy and growth



Together with a number of colleagues, I have been investigating innovation, using open data and new tools from theoretical physics and mathematics. Most economists would agree that innovation is the key driver of prosperity in advanced economies over the long run. As New Zealand is a relatively small economy that is exceptionally distant from its trading partners, we have been studying the role that scale and connectivity play in innovation. This work is now used by the Ministry of Business, Innovation and Employment to benchmark regional innovation in New Zealand.

Using open data on patents, we can examine the “economic geography” of innovation in countries, regions, and cities. Patents are a venerable source of open data: in return for the grant of a fixed-term monopoly, inventors can disclose the details of their invention so that others can build on it. Large patent datasets can now be geocoded and name-harmonised, providing rich, detailed information about the invention process itself. In the future, we hope to combine this with other data from government so that we can better track the value of innovation.

For example, if we look at the number of patents per capita in Auckland and Adelaide, two cities with similar populations, we find that the number of patents per capita turns out to be similar. However, much larger cities, like Melbourne and Sydney, produce many more patents per capita. In fact, Australia produces about 30 percent more patents per capita than New Zealand, almost entirely because of the high output of Melbourne and Sydney. This seems to be quite general – bigger cities produce more patents per capita and bigger population centres drive innovation.



New Zealand's largest inventor network on Google Earth. Inventors in New Zealand that have filed patents in the European Patent Office have been geolocated and are linked if they share a patent.

This is an important finding for New Zealand, which doesn't have a truly large city by international standards. It suggests we need to learn to collaborate as if we were a city of four million people. Indeed, there are a number of small countries, such as Denmark, Finland, and Israel that have overcome the disadvantages of size to build successful, high-technology industries with scale in areas unrelated to previous strengths.

When we look at who is connected to whom in New Zealand's largest network of inventors and at our international connections, we find that the largest group of people who have collaborated in this country is a network of about 450 inventors. These inventors work for 14 different New Zealand companies, four Crown Research Institutes (CRIs), three universities, and a private research institute. The connections are often quite surprising. For example, these links don't always seem to be sector-based.

A nice example of this is Weta Digital in Wellington. We are proud of having one of the top computer graphics companies in the world and Weta Digital's contribution to movies such as *The Rise of the Planet of the Apes* and *The Hobbit*. But what does Weta Digital want from the New Zealand science system? You might expect it wants access to ICT researchers or parallel computing. Actually, what you find is that it is working with researchers at the Auckland Bioengineering Institute, because to make a convincing animation of an ape or person, Weta needs to have an engineering model of how people and apes move.

If we had sat down 20 years ago and said we want a film industry in New Zealand, it would not have occurred to us to put public funding into Auckland's Bioengineering Institute. But today, it is one of the key players in New Zealand's innovation ecosystem that will help Weta Digital stay ahead of the game.

If New Zealand is to become as prosperous as other advanced nations we need a rich and diverse innovative ecosystem that will enable us to fill in these gaps. Because of the variety and unpredictable nature of the science we will need in the future, we must strike a balance between generating new knowledge across the board and picking winners based on past success.

Prof Shaun Hendy FRSNZ

For more information about this research, see sciblogs.co.nz/a-measure-of-science/2011/04/14/the-new-zealand-innovation-ecosystem-map/. Prof Hendy and Sir Paul Callaghan have discussed innovation in NZ in their 2013 book *Get off the Grass: Kickstarting New Zealand's innovation economy* www.amazon.com/Get-off-Grass-Kickstarting-Innovation-ebook/dp/B00FD3HSZ0/ref=tmm_kin_title_0

As Shaun's example illustrates, we think that the advantage for the country that enables this kind of research will be enormous in direct benefits to society, AND in being able to attract academic talent to New Zealand who can use this shared data to undertake world-leading research.

If researchers could use health records linked to what people eat, use, and where they live, it may be easier to find links between such things and health conditions, risks from the things in your built environment or other potential hazards. This would enable highly specific public health advice relevant to the New Zealand situation, as well as provide for world-leading research. The beauty of this is that this can be done with de-identified records and so is not an invasion of privacy.

However, if highly specific risks are identified, it should perhaps be possible to seek additional consent to find the real people who are at risk and get health checks.

Benefits to New Zealanders

The benefits described above include improvements to management of government, provide new opportunities to business, boost the capabilities of New Zealand researchers, and support improved environmental outcomes. There are already some very smart people here in New Zealand working to make significant improvements to all aspects of lives merely through better use and availability of shared data. For all New Zealanders the direct and indirect benefits include:

- new kinds of services to improve and enrich lives
- jobs in the high-paying knowledge sector
- improved cities and urban planning that will improve the places and spaces we live and work
- more effective use of resources and monitoring of pollution that will make transparent the need to improve environmental outcomes, allow communities to hold polluters to account, and reduce waste
- where government services are better planned, earlier, and targeted there will be less victims like Marc Smith and, in turn, fewer of his victims
- where government can do more with less, learn quicker, and be more accountable and transparent, this will mean either more services or less tax
- more open, more transparent, and more participatory democracy enables more people and groups to use the data to find out for themselves and mobilise their own responses to community issues.

An economically and socially prosperous New Zealand Inc.

The country that gets this right will have a competitive advantage and be more prosperous in the same way as those who embraced the Industrial Revolution over a hundred years ago. For instance:

- Where universities have access to world-class data, they will attract world-class talent and do world-leading research
- In a data-driven world, if New Zealand does data sharing well and safely, and provides a trusted environment to do so, then New Zealand has a competitive advantage in our ability to collaborate across industry, government, and academia unlike any other nation. Whether it's milk transport optimisation, traffic management, infrastructure planning, mapping food consumption to health, understanding weather and shopping behaviour, precision agriculture to reduce pollution, real-time insurance management, faster search and rescue, predicting fraud and crime, or more transparent and participatory democracy, New Zealand has a unique opportunity to play an important leading role in becoming a centre of excellence in data-driven innovation

- As Shaun Hendy here in New Zealand suggests, that kind of learning environment, with the ability to network across sectors and disciplines, would help drive innovation and usher in an era of growth and prosperity
- As Roger Denis is finding out, for the knowledge sector itself, New Zealand's small size and ability to share diverse datasets may encourage overseas investment and the ability to grow and attract talent by being the place where people can innovate with shared data, and invest in this kind of innovation safely.

Our independence, scale, and stable socio-political environment lends itself well to creating unprecedented public good and economic value from our collective data assets, without the need to infringe on privacy or data protection privileges. To make this a possibility, New Zealand needs transparent and easy-to-understand data policies, a common operating framework for data usage, and world-class data capabilities at a globally competitive scale. In creating this digital ecosystem, New Zealand will unlock opportunities for businesses, public and community service providers, government, academia, individuals – in other words, for all New Zealanders.

However, it really is a case that you must build a trusted digital eco-system to reap these benefits. It is only in a trusted data sharing environment that people will feel safe enough to share their data. Without trust in the system governing the use of shared data, it isn't really going to work. Therefore, this is only possible if we can genuinely meet some of the challenges raised in the next section, by acknowledging and having a real response to real risks around sharing and managing personal data.

3. Challenges and risks

A lot of people, when they think about the sharing and use of data, think about the US National Security Agency (NSA), the revelations of Edward Snowden on mass surveillance, and about embarrassing privacy blunders in the public and private sector. Together, these events illustrate how things can and do already go wrong with the sharing and use of personal data.

A lot of the risks in this area have been exposed through the media. Over recent months, we've seen a steady stream of stories about the privacy and security challenges around the sharing and use of data.

Case study: Bank's embarrassing privacy breach

Rob Kidd, Sunday Star Times, 8 December 2013:

www.stuff.co.nz/business/money/9490703/Banks-embarrassing-privacy-breach

IN THE MONEY: Two-year-old Joel Morrison received bank statements from a bank featuring the details of some high-profile customers. [A Bank] has been left red-faced over a major privacy breach after it inadvertently sent bank statements for customers' accounts containing hundreds of thousands of dollars to a two-year old boy. The bank botch saw Whangarei toddler Joel Morrison sent a pile of other people's account statements, all of which contained considerably more than his own savings account of about \$200. The statements arrived in the mail after Joel's mum, Stacey Morrison, requested details of her own spending. Instead her son received a statement on his account – along with those of a bunch of strangers.

When visited by the Sunday Star-Times, Joel seemed more concerned with his toy cars but his mum said it was a worrying situation.

"It's like someone coming into your house, knowing how you've spent your money; even if you've done nothing wrong," she said. . . . It's a huge concern. . . . Some of these people have hundreds of thousands."



Photo © PETER MEEGHAM/Fairfax NZ

Pushing the boundaries of privacy and people's rights as a business model

In New Zealand over the last 18 months, we've seen a series of embarrassing breaches of people's data caused either by malicious hackers cracking into data (Yahoo, Xtra, LinkedIn), by well-intentioned hackers wanting to highlight weaknesses so that these can be addressed (Work and Income kiosks), or by people mistakenly sending digitised information to the wrong address (ACC, EQC). In a paper-based environment, these breaches might have revealed one or perhaps a few names. In the world of shared digital data, however, these breaches can easily expose tens or hundreds of thousands of people's personal information.

Creepy or convenient?

Entrepreneurial global firms like Google, Facebook, LinkedIn, and Apple have been testing the limits on people's readiness to share personal information without consent with service offerings that drill into the digital personal sphere, such as street-view photography of people's front doors, text analysis on people's comments, or doing facial recognition of images without consent. Those photos you load up on Facebook and other social media sites are data-mined to determine who is in them. This is then used to link you to other people (in the photo) and learn more about your network. One reason Google wants to link your email address (and is now asking for your cell-phone number) could possibly be to build a better profile and link more disparate data sources to find out more about you.

Case study: The incredible story of how Target exposed a teen girl's pregnancy

Gus Lubin, Business Insider Australia, 17 February 2012.

www.businessinsider.com.au/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2

Target broke through to a new level of customer tracking with the help of statistical genius Andrew Pole, according to a [New York Times Magazine cover story by Charles Duhigg](#).

Pole identified 25 products that when purchased together indicate a woman is likely to be pregnant. The value of this information was that Target could send coupons to the pregnant woman at an expensive and habit-forming period of her life.

Plugged into Target's customer tracking technology, Pole's formula was a beast. Once it even exposed a teen girl's pregnancy:

[A] man walked into a Target outside Minneapolis and demanded to see the manager. He was clutching coupons that had been sent to his daughter, and he was angry, according to an employee who participated in the conversation.

"My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

The manager didn't have any idea what the man was talking about. He looked at the mailer. Sure enough, it was addressed to the man's daughter and contained advertisements for maternity clothing, nursery furniture and pictures of smiling infants. The manager apologised and then called a few days later to apologise again.

On the phone, though, the father was somewhat abashed. "I had a talk with my daughter," he said. "It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology.

We know of two companies in New Zealand that collect people's personal information and link it across diverse data sources. One with people's consent, the other "legally" triangulates using data from various public and private sources to build a profile of your household, without your consent (as far as we can see). This information is sold to insurance and other companies who then use it to sell life insurances to people. Or, in case you have been identified as a high-risk person, to try to *avoid* selling life insurance to you or (for example) people from your ethnicity. This is a case in which your personal data is potentially being used against you.

These companies are testing the limits of the law as they move into this new world of data. For them, it is big business – on-selling marketing opportunities to other companies who "target market" products to use based on the individuals' profiles they are building.

Privacy regulators and the media have countered with complaints and investigations in a number of countries. For some people, new services have moved quickly from cool to creepy, but they are still using them. To others, these services are a big turn-off. For example, some people are choosing not to use mainstream social media because they don't like the idea of their personal information being out there and available to others.

If people are data- and privacy-savvy, they can often make the system work for them. But if they are vulnerable or naïve they may be sharing their personal data in ways that make it easy to use it against them, to be duped.

Scenario: When privacy is breached?

Sure, I have a number of things that I would rather were not shared with the whole world, I guess maybe everyone does. I don't want everyone to know what I earn, what I owe, that I had mental health issues and depression as a young man, or that I am homosexual. For me, it's not been an easy life so far. It's taken me a long time to come to terms with my past and my current life, but I've got on track with the right help. I was getting by OK – until someone managed to find out. When my personal information was broadcast, all of a sudden the media picked up the story, my boss and everyone at the school where I teach thought they knew about me in ways that I had no control over. I am so ashamed that I haven't been to my classes or to work for the last few weeks. I don't know how I can show my face again. I am thinking that perhaps I'll have to move to another town, again.

What's private depends on each person's individual situation and perception. Something that seems completely innocuous to one person can be a big deal and deeply harmful to another.

How do we cater for this diversity? When mistakes are made, and personal data is released to the world, the vulnerable usually suffer most.

Example: Using the data wisely

Sometimes data-driven linking and analysing doesn't get it right. For example, in an interview he recently gave at a conference, the NSA whistleblower Edward Snowden insisted that his actions of releasing secret documents to *The Guardian* had improved the national security of the United States rather than undermined it. As an example, he pointed out that US security agencies were failing to pick up would-be terrorists, such as one of the people involved in the Boston Marathon, because they were harvesting too much data: "We are monitoring everyone's communications rather than suspects' communications," he explained. "If we hadn't spent so much on mass surveillance, if we had followed traditional patterns, we might have caught him." His message to technology firms such as Facebook and Google was "not that you can't collect any data" but rather "that you should only collect the data, and hold it as long as necessary, for the operation of your business".

www.theguardian.com/world/2014/mar/10/edward-snowden-nsa-leaks-sxsw

Well-meaning 'black hat' community and national interests are *only sometimes* in your interests

Government agencies responsible for border and national security have also come close to the boundaries of lawful investigations as they make use of freely available and shared data to pursue their purpose more effectively.

As we have seen with the NSA as an extreme case, governments want data to catch bad guys. They want this to fight crime, child abuse, terrorism, and other "national security" threats. Of course, assuming it works, this is a benefit to most people a lot of the time. We want fewer bombs, less crime, and fewer dead or abused kids. The use of profiling methods probably *will* help stop crime – by linking data and analysing it.

But sometimes linking, analysing, or even profiling doesn't get it right.

Scenario: As Mr B Dylan once said; "It ain't me babe, it ain't me you're looking for"

When I tried to visit my aunt in Sydney, I couldn't travel overseas. And my bank accounts were frozen. It turns out that my name and date of birth are an exact match with those of someone who has a criminal record and a lousy credit rating. It took a while to get it fixed with the bank, and meant I missed my aunt's birthday party. It's really frustrating when these things happen.

Misuse or poor use of data by government and powerful interests

Sometimes governments do not always work in their citizens' best interests, either through incompetence or bad behaviour. From time to time there have been allegations here in New Zealand of people misusing personal information held by government to silence dissent, such as spy agencies targeting lawful protesters.

Minority groups can be unfairly targeted – with the result of being overrepresented in the justice sector. It could be more sinister too. Powerful lobby groups and interests can drive tough laws to restrict freedom. Governments and public attitudes change. Public attitudes towards religion, ethnicity, or sexual preferences can ebb and flow – not always in the direction of more tolerance. What would you do if the politics of intolerance and the coercive powers of the state turned against your group *and* where your lifestyle is transparent to them through previously shared data?

Even if we have mechanisms to stop actual misuse of government or business data, there may still be a corrosive effect on democracy and free speech where some powerful individuals have access to your personal and very private information. The overhanging threat of potential release may affect the way citizens behave in a free society.

How do we build a good and responsible system for data sharing that can last? One that can resist the rule makers from changing the rules for dodgy reasons, yet can adapt to rapid change?

The unintended consequences of otherwise good intentions

Sometimes shared data can be done in a well-meaning fashion but have unintended consequences that can be harmful. What if a child that has unexplained bruising was too scared to tell the doctor how it happened, in fear that information would be shared with Child, Youth and Family, or the Police, or that their family would be split up?

The point here is that pervasive, unthinking data sharing may in fact lead to more harmful outcomes for some people; whether assumptions or beliefs about data sharing are true or not. Minority groups who have fraught relationships with government agencies and who just want to be left alone, may consider themselves more targeted, more invaded, and more unfairly treated, in spite of otherwise good intentions. They may be more likely to go undercover and so less likely to receive help.

What about incompetence or bad behaviour?

The best systems in the world will fail sometimes because it is inevitable that data will be lost, stolen, de-encrypted, re-identified, or used illegally. The more information is shared, the more it becomes an attractive target. Big data is big business for the criminal fraternity too who are adapting well to our digital future. What do we need to consider for when this happens? How can we detect cybercrime like bullying, extortion, identity theft? There have

been cases in New Zealand of officials selling personal data to private detectives. What repercussions and sanctions should be in place in the law? How can we compensate the victims? Does this even work trans-nationally?

Rogue governments and criminals will want to access and use shared data. Through poor data management, mismanagement, and human error, privacy breaches will happen. What then do we do?

Scenario: Losing it all

At first it seemed fantastic getting all these amazingly easy services through my smartphone. But then I tried to get into my email a few days ago, and the system wouldn't accept my password. Now I can't access anything on my computer, and my bank accounts have been emptied out. My credit cards have all been maxed to the limit and beyond – and my bank has called to tell me it's cutting me off, despite all the years I have spent conscientiously paying the mortgage. My friends and family have phoned to say they've had weird messages from me asking for them to send money. Someone has claimed my life and stolen everything from me. This is a living hell – how can I make it stop?

Identity theft is increasingly common overseas as a means to access people's bank and property details and steal their assets. The more data is available and shared, the higher the possibility of this happening. We need to put adequate frameworks in place in order to prevent this. But with more data shared and being used by more people, the risks of misuse and the stakes are much greater. Can that level of risk really be contained?

Escaping your past or merely wanting to not be annoyed by ads all the time

There is a hilarious and sobering video on YouTube from the Electronic Frontier Foundation:

www.youtube.com/watch?v=RNJI9EEcsoE

This is a wry take on the use of data sharing for target marketing and imagines a digital future where a person at a contact centre knows more about you than yourself when you merely ring to order a pizza.

There is a serious side to this; how do you move on from your history? And can you opt out of this?

"Behavioural Analytics" is an (not so) innocuous technical term for people using big data to profile your behaviour to make predictions about what you want to do or buy. This can be great in the hands of your physician who is predicting a heart attack, but what about some other uses?

We all know the past is a great predictor of the future. But it is also true that some people can rise above the tyranny of their histories. In fact, these are the heroic stories we celebrate. But how hard is this going to be when marketing companies focus on what is most likely for you, assuming you won't beat your history? Their assumption is not that you are one of those rare cases of somebody rewriting their own story, and bucking your upbringing. So what do you do when it is your past that you want to shake off?

Are you a recovering alcoholic or gambler who wants to transform your life? How will this be possible when some inspired entrepreneur will want to target these kinds of addictions via every email or billboard, or when you visit Amazon, Google, or even when some idiot is trying to sell you something. This can be merely annoying, or it can be a serious matter, affecting your will and your freedom from interference and to self-determination.

Right now, you can opt out of telemarketing by ringing up and having your phone number put on a register that means the telemarketers leave you alone. Can the same be done in a future where the information people have about you is so much more personal and the targeting more personal too?

Profiling and predicting, in detail, may create some unexpected headaches

Consider also the prediction of risk by the public or private sector. If some of the claims are true, then we will be using people's shared personal data on, for instance, eating, exercise, social history, and genetic profile to predict lifespans, heart attacks, and increased risk of a specific disease. When interests overlap, then you may get a better service, such as increased access to health care.

Of course, sometimes you may not want to know what fate has waiting in the wings for you genetically, or want to become so unduly monitoring of your risk profile (about what you eat and do) that you forget how to live. Some fears should perhaps not be nurtured by the self-interested marketing machines of business.

But what about when governments and insurance companies want to use shared data to manage their own interests? Your interests and theirs do not always overlap. Do you have a right to keep personal data away from insurance companies, for instance? An insurance company's interest is in knowing everything they can get their hands on to manage their risk of having to make a payout – the risk of losing revenue for their shareholders. If your genetic profile, health records, and social network history of base-jumping all add up to high premiums, then you may not be able to afford or even get insurance.

In summary then, there are a host of challenges and tensions for any society that wants to play in this space; the sorts of challenges that we need to consider as a nation when people come asking to have and link up your data. Challenges to safety from theft, bullying, or persecution; challenges to your autonomy and choice; challenges to freedom from interference from well-meaning (or otherwise) businesses and governments.

People talk a lot about respecting privacy, but why is it valuable? Privacy supports your autonomy to choose your own life course safely and without interference. That is what we must try to maintain if we are going to use and re-use data more widely. It would be a high price to pay to become merely the economic and social units of business and government. In fact, it would be to give up a lot of what makes us human.

We have talked a lot about privacy and consent in this section and the misuse of private data, but there are other more pragmatic questions to consider. In the next section (four), we pose some of the other challenges that need to be considered before thinking about a possible solution.

4. Challenging questions

When you read the examples of benefits, challenges, and risks that arise from the increased use of shared data, several complex questions emerge. Although we have some ideas, the Data Futures Forum does not pretend to have all the answers, but it is important that we in New Zealand consider these questions and decide how we will adapt. Consider these questions:

- *What principles should New Zealand adopt that can both meet our needs for privacy, trust, and control, yet allow for the creative and entrepreneurial benefits promised by this revolution?*

We want to ensure that there is a high trust system for data sharing, one that mitigates many of the risks laid out in section three. Yet we also need a system that is flexible and adaptable enough to allow for bright ideas and social, business, and research entrepreneurship to build a better future, i.e. without getting tangled up in unthinkingly risk-averse bureaucracy and red tape. We need to think of solutions that are sound and strong, but not brittle. What kinds of principles and solutions are they?

- *Who owns, controls, or has decision rights about data?*

Is it the collector of the data? Certainly they may have a financial interest. The person who the data is about? They certainly have an interest. In order to reap the benefits of the data revolution, it is clear that existing datasets will be re-used and new datasets will be created. But then, who owns the resulting data? The re-user? Will they be owned by the entity disclosing or collecting the data, or will they be open by default? What about collective ownership of data, such as iwi data? How are intellectual property rights arrived at from the data managed?

Who has decision rights over data? The collector? Provider (if different)? The person or entity that the data is about? If there is a data commons, who makes decisions about the New Zealand Data Commons?

- *Who is the data custodian and what are their obligations?*

Who will look after the (newly created) datasets? For instance, who is responsible for the processing and storage of the data? Where and how will data be stored, and for how long? Who will provide safeguards for data quality and data accuracy? Who is accountable

when data gets stolen? Who will have the authority to decide on those data access rights? What happens to data if the custodian gets liquidated or sold off (to another business here or overseas)? Can the liquidator on-sell the data to pay off creditors?

- *How do we protect the digital rights of every New Zealander?*

We are living in a pluralistic society with differences in cultural backgrounds and value perspectives. Some of us are also living overseas, spread all over the world and exposed to different cultures. These cultural differences influence our privacy perceptions and the types of data we are willing to share. How could we maintain our cultural diversity and be an inclusive society in which the digital rights of every New Zealander are protected? What will be the social contract for a data-driven future in New Zealand? How do we mitigate risks of vulnerable groups in our society?

- *What are we actually providing consent for, how, and by whom?*

The value of data no longer resides solely in its primary purpose. Value also resides in the re-use of data. What do you give consent to when we cannot even imagine what possible future value that data may have? Most data re-uses haven't been imagined when the data is first shared, which raises the question of how individuals can give informed consent to an unknown. Do individuals need to opt-in to an open-ended, multi-purpose arrangement? Or are there perhaps other possible arrangements for informed consent we might be able to create? Do children have digital rights to consent before a certain age? What about you, and your family's, rights when you die? Do we need digital wills?

- *Will there be digital rights to forget or delete?*

Do we need the ability as individuals to opt out in the digital age, similar to how we can decide to opt out of target marketing campaigns of telemarketers? Do we have a right to revoke our consent with the use of our personal data? How could this be arranged? Will the digital footprints and breadcrumbs you have left earlier in your digital life, such as the public posting of sensitive pictures, haunt you for the rest of your life or even beyond?

- *Will there be sovereignty over data created and shared in New Zealand? What will New Zealand's place be in the global digital world?*

How does New Zealand ensure the best outcome for New Zealanders in a global environment where digital data crosses borders? The Internet has, with a few notable exceptions, no borders and the digital world is truly global. This offers tremendous opportunities for New Zealand where our distance to markets is a major hurdle to our prosperity. In digital space the rest of the world is a keystroke away. There are major questions, even on a domestic scale about the provenance and ownership of data, but these are amplified when global sharing is considered. How do we manage off-shore situations where governments want to have access to New Zealanders' personal information; or pan-national organised criminals – both of which are trying to access New Zealand data without your consent and becoming increasingly sophisticated about it. What if the company that stores our data is owned overseas and the data is subject to

the laws of that country? We will need to figure out how we play with the data giants off-shore. Companies like Google, Amazon, and many others will hold a wealth of information about New Zealanders. This information is of significant commercial value to them, but what are the opportunities for productive collaborations that could benefit New Zealanders? Can we use the new environment to support global connections, for iwi with people living offshore, or for innovation? What would be the effect on New Zealand's prosperity in a digitally run future?

- *What is 'de-identified' data and what are the risks around re-identification? Is it realistically possible for a person to even have full anonymity in the digital age?*

There are ways to de-identify data by stripping out names, encrypting, or jittering the date of birth. But you can, with enough information about an individual, triangulate enough other information to try to re-identify them. Try this thought experiment: imagine I de-identify every commonly revealing personal detail about you like your name, address, and date of birth. Then an analyst is re-using this data to look at the average height of people in Timaru. You would probably think this was de-identified data unless you were the one person in Timaru who happened to be over seven feet walking past the window of the data analyst.

- *What about when we don't want your consent?*

There are times when governments do not want your consent. It's obvious in cases like policing and protecting children from child abuse. We do not wish to protect the privacy of some individuals.

But there are more challenging cases. What if we could use New Zealanders' personal health data to do research to save lives? Perhaps there is a life-threatening medical condition that a small number of people have. We want to profile them and compare them to other New Zealanders without the condition. But nobody wanted to opt in to share their data, though the risk to their privacy was small. When do your interests in privacy outweigh other people's interests or the collective interest? There are other real examples overseas where people's personal information has been used to track pandemic outbreaks. If such a disaster were to befall New Zealand, who would give emergency consent to open all personal data to help stop the spread of this deadly disease?

These are just some of the kinds of questions that we need to consider before leaping to any particular solution.

5. Adaptation to change

New Zealanders have some choices to make. There are four broad options, though only one looks realistic.

A. We can leave the status quo

We can use the instruments, regulations, and capabilities that we currently have, and seek to minimise risks and maximise benefits from shared data under the current regime. We will look at this option, but maintaining the status quo looks insufficient for an adaptation to this emerging future. The level of data sharing with loose control is already eroding trust and confidence. By the same token, some data sharing that would be useful is not currently sanctioned. The status quo doesn't manage risk well. Nor does it facilitate many of the benefits possible through increased data sharing.

B. Focus on privacy only and reduce the ability to share data

We can tighten up, focus mainly on the privacy risks and make it harder to share data. We think that locking data up to minimise the risks is a bit like a retrenchment policy in the face of an upcoming storm. It is not feasible or realistic to Balkanise New Zealand away from this global emerging future. Moreover, lives will be unnecessarily lost and our economy will not adapt when others do. There will be lost business opportunities and the state sector would be an opaque monastery for Wellington's mandarins. There would be continued mistrust in non-transparent government. Although there may be some merit in tightening up and waiting to see what emerges elsewhere first, we would have to forgo the current benefits we do enjoy in virtue of data already being shared usefully today.

C. Focus only on allowing increased data sharing

We might decide that these risks are not great and that the benefits far outweigh the risks involved in increased data sharing. So, open the door and not worry too much about the risks. Isn't this all just a small minority of people presenting theoretical risks? We don't think so. As we saw in section three, there are very real risks from misuse of personal data. Shared data can clearly be used to harm people. It would be foolish not to try to minimise the potential harms that arise from the use of shared data. Moreover, privacy has been around since Adam wore a fig leaf, and isn't something we should throw away lightly.

In any case, ignoring people's concerns and muscling this kind of thing in would be self-defeating. If sharing of data is based on trust, but there is no trust, then people will stop sharing it. They will vote with their feet. They will store data and use services that they know they can trust (overseas). They will emigrate. They will use paid private services (if wealthy enough). They will stop seeking help from government. Sharing data badly is not a sustainable option. From a humanistic standpoint, this would be a miserable kind of existence.

D. Try to *maximise* the benefits and *minimise* the harms of data sharing.

The Data Futures Forum thinks we can have our cake and eat it too. We want to see what steps can be taken to improve both sides of this equation. We want to consider ways that we can try to maximise the benefits through more data sharing AND improve our ability to manage and minimise the potential harms of this at the same time. We think the dialogue should be about how New Zealand can effectively adapt to reap the benefits of this future, but do so in a safe, secure, transparent, and privacy-friendly manner.

6. How can you be involved?

We want you to participate in this discussion:

- What do you think about New Zealand's possible data futures?
- What kind of benefits and opportunities should we be aiming for?
- How can we make it easy to share and use data when it is required?
- What risks and challenges need to be managed?
- Are there privacy-friendly ways to access and use data?
- How can we protect the interests of New Zealanders and New Zealand?
- What will a sustainable and adaptive environment for data sharing and use look like?

There are a number of ways you can get involved:

- Visit our website: www.nzdatafutures.org.nz.
- Join, comment, or lead a discussion on our LinkedIn discussion page: New Zealand Data Futures Forum.
- Register for the speaking events and conferences we will be engaging at – you can do this through our website.
- Invite Forum members to speak at events or discussion groups – contact us via our website.

7. The work of the Forum

Members of the Forum have until June 2014 to explore ways forward for the kinds of issues canvassed in this paper. How can we do more data sharing in New Zealand *and* build the way of doing it that builds confidence and trust and safety.

We decided at the start that the way we went about giving that advice was just as important as the advice itself. A key part of our role is to provide a forum for New Zealanders to engage with us in this debate. Between April and June 2014 the members of the Forum will be engaging with individuals and stakeholder groups, including communities, businesses, government, NGOs, M ori, and researchers to stimulate an informed debate. This does two things:

- It lets you know how we arrived at our conclusions – whether you agree with them or not. We want to be transparent.
- It helps us because we do not have all of the answers – nobody does. This is not just a technical problem. It may ultimately be a change to society and the way we live. We need your help to put up your own suggestions and your help to push back by picking holes in our suggestions. That is bound to lead to a better result.

This approach is reflected in the paper you are reading. At this point in the process, we are as much interested in ensuring as many New Zealanders as possible have an opportunity to start to grapple with these important and challenging issues that will increasingly define our digital future, as we are in arriving at “solutions”.

The opportunities and challenges

This first paper sets out the opportunities and challenges. We have had an initial look at the issues, opportunities, benefits, challenges, and risks. Have we got this right? Your feedback about this and what we have missed or overstated is important to us. You can post these yourself for all to see on one of our open forums (see next page). At the same time we will also be looking for input and advice on two further papers we will produce. These will springboard off the questions we have raised in this discussion paper and the perspectives you share with us.

Principles

Our second paper will be a first stab at some broad principles that we wish to test with you. If New Zealand is going to improve its ability to share data MORE often, AND on a safer, more trusted, and more privacy-friendly basis, then on what basis should we be doing this? Principles are about things like privacy, consent, and ownership and are general answers to the kinds of questions raised in this paper (sections 2–4). This is a bit like saying; “*what would success look like?*” A set of principles will provide us with a test. What must any *specific* solution need to look like to be a *good* solution? Does it meet our New Zealand principles on the use of shared data? If you have ideas on principles that New Zealand should adopt then please share them.

Solutions

Our third paper will consider what kinds of *specific* solutions might be adopted. How specifically can we address principles of privacy and consent, accountability, and other principles that we come up with? Already, people are sharing interesting ways of managing data; such as “digital vaults” and other ideas from New Zealanders and from overseas. Of course, not all solutions are technical. We need to think about governance and rules and enablers as well. We will share any ideas we come across and you can share your ideas too in the coming months. *And* your critique and comments on proposed solutions.

Forum Members

**John Whitehead, Chair**

CNZM. Former Secretary to Treasury.

Former Executive Director, World Bank

**Stephen England-Hall**

Chief Executive Officer, Loyalty New Zealand Limited

**James Mansell**

Director of Innovation & Strategy, Ministry of Social Development

**Paul O'Connor**

Founder, Director and Head of Research & Development, Datamine

**Joshua Feast**

CEO and co-founder, Cogito

**Miriam Lips**

Professor of e-Government, Victoria University School of Government

**John Roberts**

Director, Relationship Management, Department of Internal Affairs

**Evelyn Wareham**

GM Customer insights Analytics & Research, Statistics New Zealand

**David Wales**

Manager, Analytics and Insights, Performance Hub, Treasury