

How big data can be a force for good

Over the last decade, Stats NZ has developed a powerful tool for policymakers and researchers alike. Known as the Integrated Data Infrastructure or IDI, it draws upon data from across the government and contains de-identified information for people living in New Zealand. Since its development, 735 projects have used micro-level data from the IDI to monitor trends in our society, conduct research, identify inequality and make funding decisions. In this Insight, we investigate how the IDI helps us to do better research and make better decisions.

A trail of data

Our lives leave behind a trail of data. Our day-to-day activities, like receiving wages, filling a prescription, working for school qualifications, applying for a visa, or filling in digital and paper forms, all inform the official statistics of Aotearoa New Zealand.

Official statistics is the catchall term for data produced by government agencies. Producing them is an important role of government. This data allows policymakers and researchers to track trends in society or assess how we react to changes (e.g. What is the effect of COVID on our economy? How many people have been vaccinated?).

Because of this, successive governments place a high emphasis on data. In 2011, the government made the Declaration on Open and Transparent Government. In 2017, New Zealand joined the International Open Data Charter, signalling our commitment to international standards. And this month, the Data and Statistics Act 2022 came into effect. The new Data and Statistics Act 2022 supports a well-functioning government data system that protects private information. The new Act modernises the way statistics are collected, managed and used to retain public trust and confidence.

With privacy protections in place, these commitments are important if we are to get the most from these data sets

Built from the data created by our interactions with the government, the IDI draws upon data from

across government agencies. Because it operates at the individual level, data in the IDI is de-identified to prohibit the identification of individuals and protect privacy.

Although it is possible to request de-identified government administrative data directly from the relevant government agency, the acquired data cannot be linked to other administrative datasets. The IDI adds value to the government administrative datasets by linking individuals across government agencies. For example, health data acquired from the Ministry of Health cannot be linked with income data from IRD without using the IDI.

As such, the IDI is a powerful tool for researchers, who can use it to answer questions like: Is there a link between wellbeing and traffic accidents? What are the socioeconomic outcomes of cochlear implantations? What are the habits that impact oral health? All of these are research topics that have used the IDI.

Administrative data is good for research and refining policy...

In some ways, administrative data is a researcher's dream. You can see whether people are enrolling in tertiary education, have court appearances, or are showing up in the emergency department. These actual events can be far more useful to a researcher than surveys of stated intentions and preferences.

Of course, when we combine all this individual data, we build a database at the population level. By recording information on everyone, we don't need to worry about selection bias. We don't run into the same problems as results gleaned from survey data. Administrative data covers everyone.

...yet the needs of administrators and researchers do not always align

At the same time, what's useful for administrative purposes isn't always what a researcher needs.

We may know how many vehicles a person owns, but we don't know how often they drive or how much they want better public transport. Some survey data in the IDI (including the General Social Survey) includes wellbeing questions; however, the focus remains on interactions with government services. While some research questions can be safely answered quantitatively (say, what's the return on investment from tertiary education?), others require qualitative input (say, what drives hate crime?).

Although the population-wide coverage of the IDI is a plus, we can't expect data at this scale to be perfect. Government agencies are legally required to have different identifiers for people, so people are matched by Stats NZ using names and other identifying information. This matching process gets it right almost always – but it isn't perfect. Fortunately, right almost all the time is typically fine. Wrong matches here and there are unlikely to affect research results too much. Still, there is always the possibility of errors. And when research is being used to inform policy decisions, we need to know if it is reliable.

The IDI is used to improve our lives free of commercial imperatives

Most of us are already aware that businesses collect our data. Businesses can accumulate massive

amounts of data by following us on the web. Big players like Google and Facebook are experts at this. Our name, gender, IP address, engagement scores, and purchase histories, to name a few, are all collected.

With this data, companies can analyse us. Some use it to improve customer experiences, boost engagement, keep us on their website for just a little longer, or even figure out that last little push for us to make a purchase. Other companies collect our data and then sell it as a revenue source. Data brokers (Google and Facebook) find our data immensely valuable. Although they sometimes can use our data to improve our wellbeing, they are more interested in the commercial value behind our data. Unlike data brokers, the IDI is not for or for commercial use. It is a tool used only to improve the lives of all New Zealanders. It guides decisions, generates knowledge, and provides insight into solving problems, all in the name of public service.

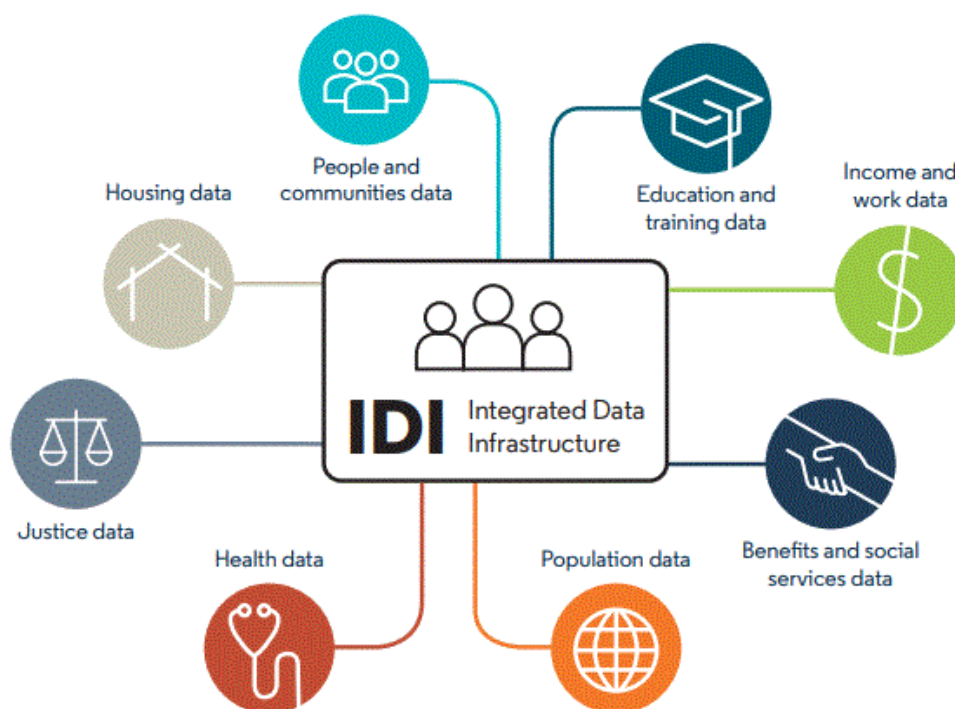
735 projects in a decade

New Zealand is one of the few countries worldwide that established an integrated data system. Stats NZ created the IDI in 2013 and has been curating and managing it since then. It is now a system containing microdata about millions of individuals living or having lived in New Zealand, fed mostly by government administrative data from health, justice, labour market, social development, housing and education.

Since its establishment, it has gained popularity amongst government organisations and private sector researchers (mainly academic researchers), despite COVID limiting access to the database.

With 735 projects approved since its development, the popularity of integrated data isn't surprising. The IDI holds anonymised personal data from almost everyone stretching back as far as 1840 in some datasets. (IDI data on births, deaths and marriages goes back to 1840 (Stats NZ 2022).) There's huge potential to generate a wide range of insights for researchers, provided they can access it.

Figure 1 Data in the IDI



Source: Stats NZ

Balancing access vs protection

Even after gaining access, access to the IDI is not unlimited. Together with the Ngā Tikanga Paihere, protection of the IDI is ensured through the “Five Safes Framework”.¹ This is designed to check that the people, projects, settings, data, and output connected with the IDI promote the values and adhere to the responsibilities of Stats NZ (Stats NZ 2020).

This restriction on access is required to keep New Zealanders safe. Ultimately, it is a balancing act for Stats NZ. We will address the issue of social licence in using government-collected personal data in a follow-up Insight.

What types of research are conducted in the IDI?

So, how is the IDI used? What research subjects are currently being investigated? And most importantly, how does this research contribute to national and individual wellbeing?

According to Stats NZ, there are currently 735 external research projects that have used the integrated data, covering a wide range of topics. We include a handful of examples here.

Out of the 735 research papers, a significant fraction are papers in the area of health. This research helps to better our understanding of various aspects of people’s health in New Zealand.

Health data in the IDI makes it possible to monitor health service utilisation over time and across geographical regions. It also enables the

¹ <https://www.stats.govt.nz/integrated-data/how-we-keep-integrated-data-safe/>

identification of inefficiencies in our health system. For example, what factors contribute to diabetes diagnosis, and how should we intervene to stop pre-diabetes from progressing to diabetes? (Teng et al. 2019). How are mental health services being used in New Zealand, and what are our unmet needs? (Gibb and Cunningham 2018).

The linking mechanism in the IDI also allows researchers to combine data from different government agencies. For example, by combining data from the Ministry of Health and IRD, researchers can look at how the diagnosis of long-term conditions like diabetes can affect people's income and employment (Dixon 2015). In 2022 NZIER extended Dixon's (2015) research and used the IDI to study the effect of prostate cancer on people's income and employment. We also compared the effect of an early prostate cancer diagnosis with late stage prostate cancer diagnosis. We found that, on average, men who get a late diagnosis lose out on about \$12,000 over four years (Hosking 2022).

Another major area for IDI research is education. This research allows us to look deeper into various aspects of our educational system in New Zealand. Examples of completed research in this area include: What contributes to tertiary education participation and achievement? (Early 2018) and What is the return on education for New Zealanders (Scott 2020; Ali 2019; Mahoney, Park, and Smyth 2013).

Of course, the IDI is also used in other research fields like income, housing, employment, tax, environment etc. (Maré and Hyslop 2021) used the IDI data to study how the minimum wage policy has changed in the past decades and how New Zealanders are affected by the change in minimum wages. In 2022, they looked at how the accommodation supplement policy has changed since 2018 and whether increased housing supplement helps to make housing more affordable ((Hyslop and Maré 2022). While (Riggs and Mitchel 2021) used IDI data to simulate how different options could achieve net zero emissions and how this can affect employment in New Zealand.

Addressing inequality

When equity features in government policy, such as our new health legislation, the IDI provides a good

tool for better understanding who gets what and to what outcome. Identifying the problem is the first step. Addressing inequality issues requires us to see how different groups interact with services, such as in our healthcare system. We also need better data to help us design effective policies which address inequality issues. The IDI provides a useful tool for us to research inequality.

For example, The Medical Council of New Zealand partnered with Te Ohu Rata O Aotearoa (Te ORA) and used IDI data to investigate patient outcomes between non-Māori and Māori. They focused on Māori patients' experiences in New Zealand, finding that Māori lagged on many of the health outcome indicators of non-Māori (The Medical Council of New Zealand and Te Ohu Rata O Aotearoa 2020).

The New Zealand Treasury used the IDI to examine the heterogeneity in mental health across different demographic groups in New Zealand and studied how different groups of people access mental health services (Brown 2019). They found that Māori have higher prevalence rates for mental health conditions, while Pacific and Asian people are less likely to have a mental health prescription or referral.

The IDI also made it possible to investigate other aspects of inequality in New Zealand, including the wage gap between genders and ethnicities.

Using data from the IDI, Sin, Stillman, and Fabling (2017) estimated that the gender wage gap increases with age and tenure when comparing men and women with similar roles and productivity.

By estimating the ethnic wage gaps in New Zealand, the inequality research also paves the way for understanding inequality in our system. For instance, The Treasury (2018) estimated education qualifications account for 18–22 percent of the Māori-Pākehā wage gap for males and 22–25 percent of the Māori-Pākehā wage gap for females. This implies that effective policies encouraging education attainment across ethnic groups may reduce wage inequality.

Fighting COVID-19

The IDI has also been used as a powerful tool in the fight against COVID-19.

For the public sector, the IDI is a good source of information to monitor long-term vaccine uptake and government benefit payments. The Social Wellbeing Agency used the IDI to research COVID-19 vaccinations among disability groups (Social Wellbeing Agency 2022).

The Ministry of Social Development has been using the IDI to monitor how businesses and people use COVID-19 income subsidies across the different income support schemes (Ministry of Social Development 2022).

Other organisations are using the IDI to analyse the effectiveness of the government's COVID-19 response in New Zealand.

Harvey et al. (2021) used IDI data to simulate the success rate for Alert Level 2.5 in preventing the spread of COVID-19 in the community. They found that Level 2.5 has a higher chance of quelling the spread of the virus if the initial outbreak is small but would be insufficient in stopping the virus within 150 days of detection if the initial outbreak is large.

Hyslop and Maré (2021) analysed the effect of the government's COVID Wage Subsidy on labour market flows, and they found a large drop in job turnover rates for firms that received the subsidy. Pacheco, Plum, and Tran (2022) investigated how COVID-19 affected the disparities between Pasifika and Pākehā in the labour market and found that although there were significant pre-pandemic ethnic disparities, the pandemic amplified these.

Who is doing the research?

Access to the IDI is not limited to government organisations. Although government administrative datasets make up most of the data in the IDI, the use of the datasets is available to the public and private sectors alike.

The University of Auckland has pulled together a comprehensive list of research outputs from IDI projects and classified them into Government reports, Non-government reports, journal articles, non-empirical reports and theses. From the list of research papers they compiled, 229 papers were prepared by various government organisations, while 146 papers were produced by non-

government researchers, including consultancies and university academics.²

The IDI is used for policy decision-making

It's clear that the IDI is a very useful tool for research – but how is it used for policy decisions? What are the tangible payoffs? Are investments into the database affecting how policymakers think and act? And what are some of the decisions our government have made based on IDI research?

The IDI is being used to help the government prioritise its funding decisions. This is often done by researching and estimating the effectiveness of a certain policy after it has been implemented. Policymakers can then decide whether to cancel, continue or expand the scope of the policy based on real-world evidence. Here are some examples of that process.

Twinking programme delivery

In 2001, New Zealand launched its Individual Placement and Support (IPS) approach to help people with mental health issues acquire and hold employment. This programme is designed to offer tailored support to individuals with mental illness. The programme is largely funded by local health budgets rather than the central government. This funding pathway makes it easier to access IPS in some regions but more difficult in others, which promotes inequality across New Zealand (Work Counts 2022). Although the funding for IPS has gradually increased and was scaled up until 2017, the programme remained relatively underfunded, with low and uneven access across District Health Boards. As of 2019, there are only around 3.7 full-time equivalent IPS specialists per 10,000 people with mental health and addiction issues (Ministry of Social Development 2020).

² The list of research projects can be found here: https://www.zotero.org/groups/4681141/idi_publications/library

Figure 2 IPS services in New Zealand

Consultancy services	
Fidelity reviews	
Employment support training	
Bespoke training	
Tools and templates	
Practical implementation support	

Source: Work Counts

In 2020, the Ministry of Social Development published a report reviewing the employment outcomes for IPS participants using the IDI. By linking mental health data, benefits, IRD data on wage and salary, and Corrections data on sentences served, researchers could compare the employment outcomes of IPS participants with those outside the program. They found that people with mental health conditions tend to be highly disadvantaged in the job market and that IPS participants experience significantly better employment outcomes (Ministry of Social Development 2020).

This IDI research shows how IPS is an effective programme supporting employment, especially for highly disadvantaged people in the labour market. Based on these findings, in 2022, the Ministry of Social Development and the Ministry of Health started working together to increase funding for the IPS and access to this programme (Work Counts 2022).

Using evidence to improve effectiveness

The Youth Service for those not in Employment Education or Training (YS:NEET) is a government-funded programme that helps 16–18-year-olds to obtain qualifications and skills. Under this program, the Ministry of Social Development helps fund social services, which help ease the transition of young people without qualifications or certified skills into the labour market.

Since its inception, the program has received adequate funding and has been popular among local service providers. However, IDI data research has shown the programme did not target disadvantaged youth very well. It also did not deliver any increase in the employment rates of

participants nor a decrease in the rates of benefit uptake. This is just one example of how even a well-funded and widely used programme can be improved with evidence-based analysis. As a result of IDI research, the service has been adapted to include more training and resourcing schemes (Light 2020).

Young people who are struggling to reach their potential can benefit immensely from evidence-based interventions. The IDI is one data source that can greatly improve services affecting our young people.

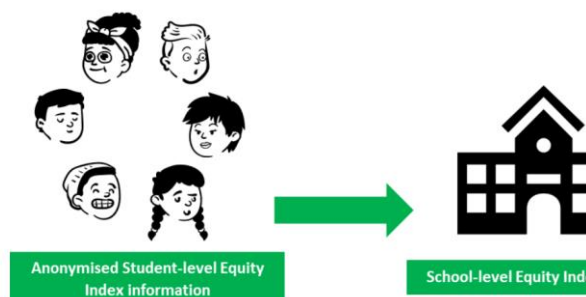
Reducing inequality – The Equity Index

The IDI is also used to reflect better what we know about our society. Allowing us to better monitor the system we operate and help those in need.

Undoubtedly, there are inequalities in our education system, from the unequal distribution of academic resources to the social-economic status of students during their education. To address this inequality, our government pours more funding into those schools that are relatively disadvantaged compared to others. Currently, the socioeconomic position of a school is measured using a decile system, which is re-calculated every five years using the Census data (Ministry of Education 2015). This has been a very long-standing practice, with the school decile system reaching back 30 years. However, this has been criticised for being too blunt and flawed, basing calculations on neighbourhoods instead of the students themselves (Dubby 2022).

In 2019, Cabinet announced they were removing the current school decile system and replacing it with the Equity Index. Powered by administrative data in the IDI, the Equity Index is designed to better reflect the disadvantages students face in New Zealand schools. It's calculated based on 37 socioeconomic factors, which are shown to affect educational outcomes. The 37 factors include a wide range of measures, including parents' education level to benefit history. These factors are then turned into weights to calculate the Equity Index. It is believed that using the Equity Index will allow for more targeted government subsidies to schools that need financial support the most (Dubby 2022). From 2023, the Equity Index system will officially be used to distribute funding (Ministry of Education 2019).

Figure 3 Calculation of the Equity Index is based on student-level information



Source: Ministry of Education

What next

The IDI has been a quiet success in its first decade. With 735 projects approved, the IDI has improved

our lives in tangible ways. Benefits range from better data, conducting research and monitoring results, addressing inequality, and making fairer funding decisions. The powerful structure of the dataset makes it an ideal tool for research in New Zealand, and research output has been growing significantly in the past decade.

Despite the various benefits of the IDI, the existence of the database is still, in many ways, unknown to the general public. Like many organisations, public trust is vital to Stats NZ's operation. Given the vast amount of individual-level data in the IDI, this database's existence needs to retain public trust, and this requires purposeful management.

Maintaining a social licence for the IDI to operate will be the subject of a follow-up Insight.

References

- Ali, Asaad Ismail. 2019. "Returns to Initial Years of Formal Education: How Birthdate Affects Later Educational Outcomes." <https://doi.org/10.26021/5333>.
- Brown, Simon. 2019. "Wellbeing and Mental Health: An Analysis Based on the Treasury's Living Standards Framework." Report. New Zealand Treasury. New Zealand. <https://apo.org.au/node/246941>.
- Dixon, Sylvia. 2015. "The Employment and Income Effects of Eight Chronic and Acute Health Conditions (WP 15/15)." The New Zealand Treasury. 2015. <https://www.treasury.govt.nz/publications/wp/employment-and-income-effects-eight-chronic-and-acute-health-conditions-wp-15-15-html>.
- Dubby, Henry. 2022. "Dumping Deciles: What's the Equity Index and How Could It Affect Your School?" *NZ Herald*, 2022, sec. New Zealand, Education. <https://www.nzherald.co.nz/nz/dumping-deciles-whats-the-equity-index-and-how-could-it-affect-your-school/U6LNNTBORV7QGA55XRSR56XHME/>.
- Early, David. 2018. "Going on to, and Achieving in, Higher-Level Tertiary Education." Ministry of Education. 2018. <https://www.educationcounts.govt.nz/publications/80898/going-on-to,-and-achieving-in,-higher-level-tertiary-education>.
- Gibb, Sheree, and Ruth Cunningham. 2018. "Recent Trends in Service Use, Unmet Need, and Information Gaps," 77. <https://www.ecald.com/news-and-updates/news/mental-health-and-addiction-in-aotearoa-new-zealand-recent-trends-in-service-use-unmet-need-and-information-gaps-2018/>
- Harvey, Emily, Oliver Maclaren, Dion O'Neale, Frankie Patten-Elliott, and David Wu. 2021. "Alert Level 2.5 Is Insufficient for Suppression or Elimination of COVID-19 Community Outbreak," 14.
- Hosking, Mike, dir. 2022. "Research Finds Late Prostate Cancer Diagnoses Cost NZ \$300 Million over Four Years." *ZB Newstalk*. <https://www.newstalkzb.co.nz/on-air/mike-hosking-breakfast/audio/sarah-hogan-nzier-economist-on-research-finding-late-prostate-cancer-diagnoses-cost-nz-300-million-over-four-years/>.
- Hyslop, Dean, and David C Maré. 2021. "Covid-19 Wage Subsidy Support and Effects," 21.
- . 2022. "The Impact of the 2018 Families Package Accommodation Supplement Area Changes on Housing Outcomes," 100.
- Light, Rowan. 2020. "Catching the Tide | New Directions for Youth NEET Policy after COVID-19." *Maxim Institute* (blog). 2020. <https://www.maxim.org.nz/article/catching-the-tide/>.
- Mahoney, Paul, Zaneta Park, and Roger Smyth. 2013. "Moving on up: What Young People Earn after Their Tertiary Education." Ministry of Education. 2013. <https://www.educationcounts.govt.nz/publications/80898/115410>.

- Maré, David C, and Dean R. Hyslop. 2021. "Minimum Wages in New Zealand: Policy and Practice in the 21st Century," 79. <https://www.motu.nz/our-research/population-and-labour/firm-performance-and-labour-dynamics/minimum-wages-nz-policy-and-practice-21st-century/>
- Ministry of Education. 2015. "School Deciles." Education in New Zealand. July 1, 2015. <https://www.education.govt.nz/school/funding-and-financials/resourcing/operational-funding/school-decile-ratings/>.
- . 2019. "Equity Index." Education in New Zealand. August 5, 2019. <https://www.education.govt.nz/our-work/information-releases/issue-specific-releases/equity/>.
- Ministry of Social Development. 2020. "Individual Placement and Support (IPS) Trials - Ministry of Social Development." MSD. <https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/research/individual-placement-and-support-trials/index.html>.
- . 2022. "Who Received the COVID-19 Income Relief Payment," 6.
- Pacheco, Gail, Alexander Plum, and Linda Tran. 2022. "The Pacific Workforce and the Impact of COVID-19," 56. <https://workresearch.aut.ac.nz/document-library/project-reports/latest-project-reports/the-pacific-workforce-and-the-impact-of-covid-19>
- Riggs, Lynn, and Livvy Mitchel. 2021. "Predicted Distributional Impacts of Climate Change Policy on Employment." <https://www.motu.nz/our-research/environment-and-resources/emission-mitigation/impacts-climate-change-policy-employment/>.
- Scott, David. 2020. "Education and Earnings, a New Zealand Update | Education Counts." 2020. <https://www.educationcounts.govt.nz/publications/80898/education-and-earnings>.
- Sin, Isabelle, Steven Stillman, and Richard Fabling. 2017. "What Drives the Gender Wage Gap? Examining the Roles of Sorting, Productivity Differences, and Discrimination." Motu. 2017. <https://www.motu.nz/our-research/population-and-labour/individual-and-group-outcomes/what-drives-the-gender-wage-gap-examining-the-roles-of-sorting-productivity-differences-and-discrimination-2/>.
- Social Wellbeing Agency. 2022. "Disabled People's COVID-19 Vaccinations Reach 90 Percent | Social Wellbeing Agency." 2022. <https://swa.govt.nz/news/disabled-peoples-covid-19-vaccinations-reach-90-percent/>.
- Stats NZ. 2020. "How to Apply Ngā Tikanga Paihere to Microdata Research Projects | Stats NZ." 2020. <https://www.stats.govt.nz/integrated-data/apply-to-use-microdata-for-research/how-to-apply-nga-tikanga-paihere-to-microdata-research-projects/>.
- . 2022. "Data in the IDI." 2022. <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/data-in-the-idi/>.
- Teng, Andrea, Tony Blakely, Nina Scott, Rawiri Jansen, Bridgette Masters-Awatere, Jeremy Krebs, and John Oetzel. 2019. "What Protects against Pre-Diabetes Progressing to Diabetes? Observational Study of Integrated Health and Social Data." *Diabetes Research and Clinical Practice* 148 (February): 119–29. <https://doi.org/10.1016/j.diabres.2018.12.003>.
- The Medical Council of New Zealand, and Te Ohu Rata O Aotearoa. 2020. "New Report on Cultural Safety and Health Equity for Māori." Medical Council. September 25, 2020. <https://www.mcnz.org.nz/about-us/news-and-updates/new-report-on-cultural-safety-and-health-equity-for-maori/>.
- The Treasury. 2018. "Statistical Analysis of Ethnic Wage Gaps in New Zealand (AP 18/03)." September 26, 2018. <https://www.treasury.govt.nz/publications/ap/ap-18-03>.
- Work Counts. 2022. "Enabling Access to IPS Employment Support in Aotearoa New Zealand in 2022 and Beyond." *Work Counts* (blog). March 18, 2022. <https://www.workcounts.co.nz/enabling-access-to-ips-employment-support-in-aotearoa-in-2022/>.

This Insight was written by Dr Zhongchen Song at NZIER, October 2022.

For further information please contact Dr Zhongchen Song at zhongchen.song@nzier.org.nz or 02102218502

The author is grateful to Samuel Strickland, Stats NZ, Sarah Hogan, Dr Olivia Wills, Todd Kriebel, Dr John Yeabsley and Sarah Spring, for their advice and feedback on drafts of this Insight.

How to cite this document:

Zhongchen, S. 2022. "How big data can be a force for good ". NZIER Insight 103. Available at www.nzier.org.nz

NZIER | (04) 472 1880 | econ@nzier.org.nz | PO Box 3479 Wellington

NZIER Insights are short notes designed to stimulate discussion on topical issues or to illustrate frameworks available for analysing economic problems. They are produced by NZIER as part of its self-funded Public Good research programme. NZIER is an independent non-profit organisation, founded in 1958, that uses applied economic analysis to provide business and policy advice to clients in the public and private sectors.

While NZIER will use all reasonable endeavours in undertaking contract research and producing reports to ensure the information is as accurate as practicable, the Institute, its contributors, employees, and Board shall not be liable (whether in contract, tort (including negligence), equity or on any other basis) for any loss or damage sustained by any person relying on such work whatever the cause of such loss or damage.