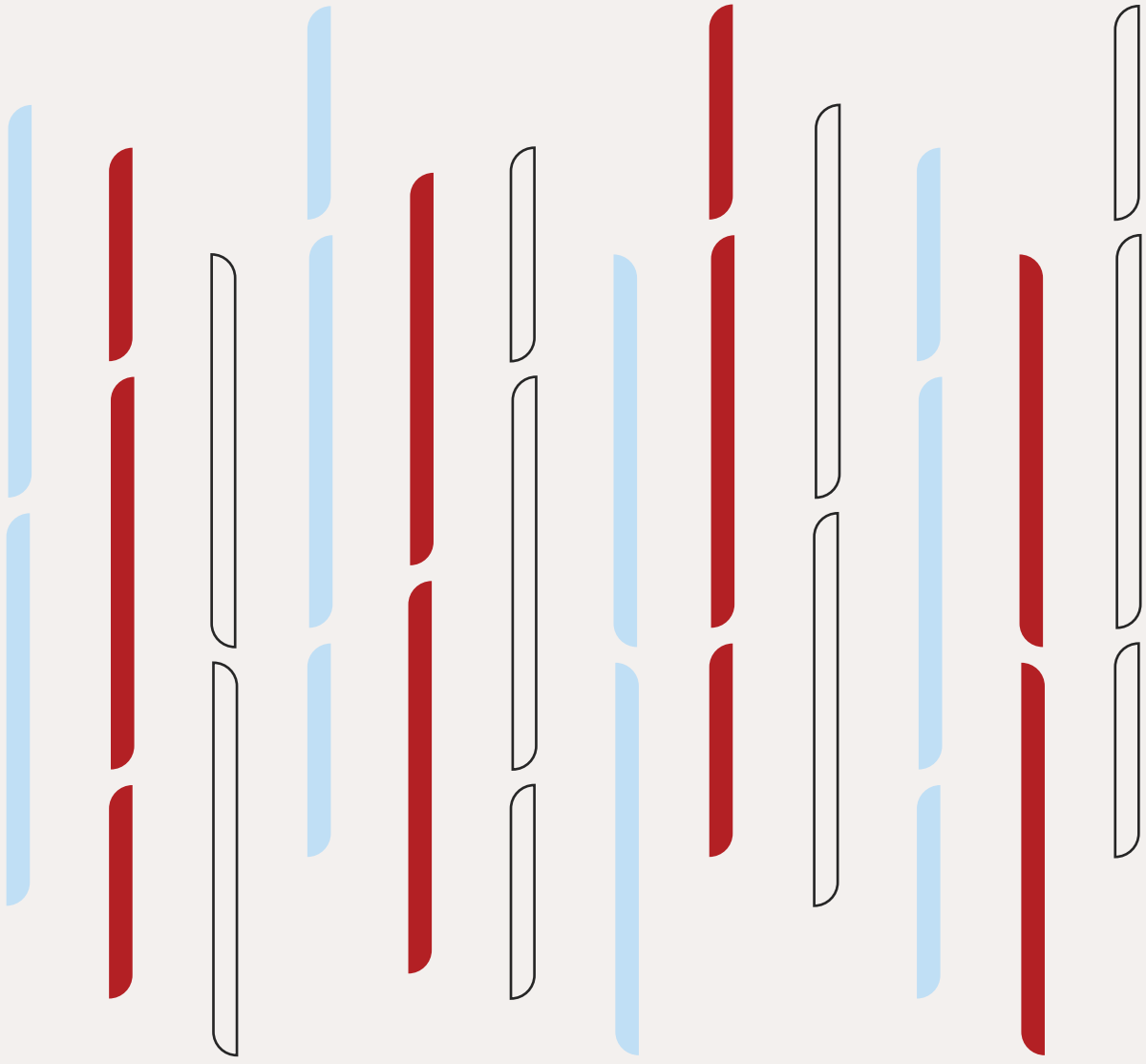
Understanding Policing Delivery

Data SaGA: Data Stocktake and Gaps Analysis

UPD
UNDERSTANDING POLICING DELIVERY

August 2024
Dr Paul Brown



Understanding Policing Delivery

Understanding Policing Delivery is an independent research programme looking at fair and equitable policing for Māori and other communities.

Both the Articles and the Principles of Te Tiriti o Waitangi serve as foundational to the programme, along with the values of Kaitiakitanga, Manaakitanga, Whakamana, Whanaungatanga, and Aroha ki te Tangata.

In the context of Understanding Policing Delivery, whanaungatanga has driven our way of working. Embodied as the creation and maintenance of strong relationships between the different rōpū who have embarked on this journey of work together

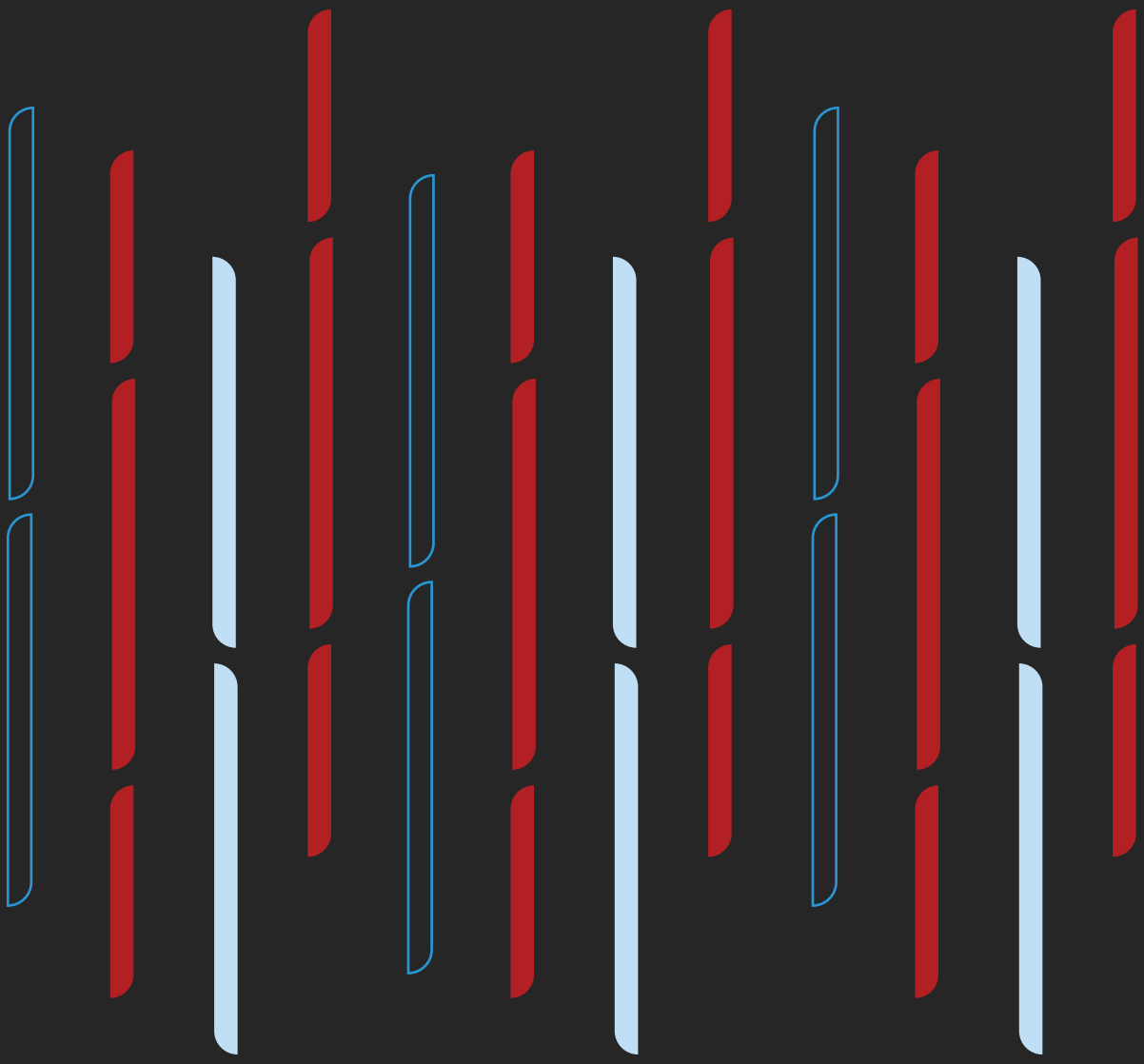
With contributions from the UPD Operational Advisory Group and UPD Ethics Committee.

Contents

Acknowledgements

I would sincerely like to thank all the interviewees that gave their time to participate in this research. I would like to thank the UPD panel for their participation and support, and the UPD management team for their hard work to help organise and set up the interviews and literature to review. Thank you to the UPD research teams, Ihi Research, Mana Pounamu Consulting, and the Donald Beasley Institute for their participation and support of this project. Thank you also to those that took the time to review this report. Your comments, feedback, and constructive criticisms helped improve the quality of the report.

I appreciate the involvement that New Zealand Police played by opening their data processes and infrastructure to our scrutiny, thank you for being supportive of this project and for being an active participant. Lastly, I would especially like to thank the Police staff who work with the data and the challenges that come along with that work. Although the challenges regarding data are immense, I came away from the interviews full of admiration for the commitment those on the ground have for improving and helping Police make Aotearoa New Zealand a safer place for all.



Introduction

Data plays a key role in modern policing. Data has a range of functions for Police, from understanding trends and patterns of crime, to allocation and deployment of resources. Police have been data collectors and users throughout their existence, from recording offences on paper, storing them in filing cabinets, and using pins on a map to gauge criminal trends.

Now, Police are using large scale databases to store data, and using sophisticated statistical methods for predictive analytics in space and time, among many other uses of data. The aim of Ngā Pirihimana o Aotearoa/New Zealand Police (NZP) is to make New Zealand the safest country, ensuring everyone can be and feel safe by preventing crime and harm through exceptional policing[1].

Police Legitimacy' is a principle that supports the idea that Police have permission to exercise authority to maintain social order, where an important component of this principle is 'Policing by Consent - to have the trust and confidence of all'. Due to the reliance of data in supporting decision-making, it is vital that data practices within Police are robust and transparent. This ensures that data is used to maximum effect – supporting good decision-making, leading to fair and equitable outcomes which enables Police to reinforce trust and accountability to the public. Apart from people, data is the most important asset that Police have!

Over the past decade, there have been significant efforts from NZP to improve their information and communication technologies (ICT) systems and data capabilities. The Police National Intelligence Centre (NIC) was established in 2009/10 and developed a consistent set of national standards for the collection, analysis, and dissemination of information. Prior to the establishment of the NIC, each station/area/district generally had their own approach to the collection and use of data. The Information and Systems Strategy and Roadmap 2013-2018 aimed to provide a clear and strategic direction for the development of information, processes, and ICT systems to ensure officers have accessible, timely, and relevant information to deliver services and be better placed to prevent crime.

The Evidence Based Policing Centre (EBPC) was established in 2017 to help better inform Police practice, implement preventative strategies, and improve the allocation of policing resources using a mix of data, predictive analytics and crime theory. The NZP ICT Strategy 2017 was developed to further leverage Police data and ICT systems to connect, inform, and enable Police in their service delivery.

1. NZP "Our Business". https://www.police.govt.nz/sites/default/files/about-us/about-nz-police/our-business-2020.jpg

Currently, the NZP Data Strategy outlines the goals, requirements, principles, and a roadmap for data and data systems moving forward. The main aspiration of NZP with respect to data is to “deliver exceptional policing services by unlocking insights in our data assets”.

The past decade of data developments has yielded some successes for operational initiatives, such as the management of COVID, policing the parliamentary protests in 2021, and the quality collection of prosecution data that has enabled statistical modelling of the likelihood of prosecution. For policy and research purposes however, the effective and efficient use of data remains largely elusive.

UPD and Data SaGA Kaupapa

The overall aim of New Zealand (NZ) Police is to deliver policing services that are fair, impartial, ethical, and just in accordance with the principles of 'police legitimacy' and 'policing by consent'.

As such, the Understanding Policing Delivery (UPD) programme aims to understand whether, where, and to what extent systemic bias may exist in NZ policing services and their overall structure. The findings of this programme will be used to develop and ensure police policy is fair and equitable to all.

The UPD programme looks at three key research areas:

- Who Police stop and speak to, and how police engage with them,
- Police decision-making around the use of force,
- Police decision-making around prosecutions.

The research will explore issues of fairness, equity, and bias with respect to many at-risk communities. However, Police-Māori relations is a major focus of the UPD.

The Data Stocktake and Gaps Analysis (Data SaGA), as part of the UPD project, aims to understand the current state of the NZP's data ecosystem from a fairness and equity lens, and investigate what data is held that is practically useful, and what gaps exist in the data, and the systems that hold the data.

There are three research questions that we attempt to answer in this work:

1. What are the important questions we wish to answer with respect to fair and equitable police service delivery, that could/should be answered using data?
2. What is the present state of the NZ Police data ecosystem, and what data does it contain?
3. What gaps exist in the data ecosystem corresponding to the important questions in 1.

The purpose is to understand what the gaps in the data are, and how we can plug those gaps through a set of recommendations.

As the author is outside and independent of the NZP, we conducted interviews with those inside of Police, data practitioners, frontline officers, managers, and groups to understand their perspectives of what data the systems held, and what the issues are regarding data.

We consulted others from outside of NZP that work with police data through the NZ Integrated Data Infrastructure (IDI), Police researchers, and Indigenous data sovereignty experts. We also consulted literature regarding the NZP Data Strategy, the National Recording Standards (NRS), and documents regarding various initiatives that have, or is currently being undertaken within Police, such as the Gender and Ethnicity Project, ReFrame, and the Disability Data Project and Roadmap.

This document reports the findings from these consultations and provides a list of recommendations to help remediate data gaps.

Data and Data Gaps

NZP data is primarily collected by front-line officers and staff in the 111 and 105 services. Data resulting from Police interactions are entered and stored in the appropriate database. There are three large databases that are managed by NZP, the National Intelligence Application (NIA), the Communications and Resource Deployment (CARD), and the Police Infringement Processing System (PIPS).

Several smaller databases also exist, such as the Tactical Operations Reporting database. From an operational perspective, Police collect core data such as location, time, identity, crime type, etc. as a function of their services. The data collected is dependent on the incident or offence that occurs, for which the general rules and standards for demographic data are found in the National Recording Standards (NRS)[2] document.

Prior to the start of this study, several briefings took place between UPD staff and data practitioners and experts within NZP. These conversations revealed that for the key research areas of UPD, the quality of data was limited, and there were large holes in the data that is needed to perform robust data analysis and statistics to gain insights into various services. For prosecutions, data was sufficient to perform data analysis and modelling as there was time for frontline staff to collect data about an offender.

Although datasets are relatively complete, there is data and information that cannot be captured in many instances that may provide greater insights into the decision to prosecute, such as the nature of the interaction Police had with the offender, or the characteristics of the officer(s) that performed the arrest. For use-of-force, data regarding tactical options are recorded and staff submit a report which goes through a review process. However, records are written retrospectively and from the officer's perspective, thus there may be bias in the reports. Structured data fields and variables also suffer from a lack of demographic information regarding those involved in the incident.

For many other interactions with Police, there are practical limitations to what frontline officers can collect. For example, for roadside stops, demographic information about the driver and passengers are not recorded. Those operating 111 calls in an emergency do not have the time to collect information other than what is necessary to dispatch the necessary resources to the caller.

We provide the reader with some context around data "gaps" as this term is broad and quite vague. As we have moved through the process, we have attempted to categorise different types of data gaps. We note here that we looked at this more from a policy/research perspective rather than an operational perspective, though the gaps affect both missions.

2. The latest version of the NRS can be found at: https://www.police.govt.nz/sites/default/files/publications/national-recording-standard-nrs.pdf. Chapters 4 and 6 summarise what Police record, and the definitions of those variables.

We have focused on the following data gaps in the context of Police data:

1. Data that is not collected well.

- For example, the ethnicity of someone interacting with Police may be collected under some circumstances, but only one ethnicity is available to be chosen. In the case of Māori, there is the ability to input two iwi with two corresponding hapū, but no more.

2. Data that is not collected at all.

- For example, in proceedings data, there is little to no data collected if an offender has any disabilities or mental health impairments. During roadside stops, demographic data is not usually captured.

3. Data that is collected but lacks quality, validity, and reliability standards.

- There are many data fields where there are errors due to data entry (e.g., dates put into locations, or the location of an incident given as the police station). These can be corrected if found, but in some instances, they cannot be corrected if the initial information is lost.

4. Data with classes that are poorly defined.

- For example, "sex" has three levels, "Male", "Female", and "Unknown". There is no option for those that identify outside of the gender binary.

5. Data that is not optimised for efficient re-use[3].

- Data is generally collected for operational purposes, thus using the data outside this purpose would be classed as secondary use. Data that is not findable, accessible, and are kept within systems that lack interoperability (the ability to join datasets from different sources) with internal systems are not optimised for re-use.

Through our interviews and conversations, we have found numerous examples of Police data that fall under one or more of these categories. We present these in the summary of findings.

3. The FAIR principles speak to optimal reuse of secondary data. See https://www.go-fair.org/fair-principles/.

NZP Data Strategy 2021-2024

The current NZP Data Strategy aspires to "deliver exceptional policing services by unlocking the insights of our data and information assets". Furthermore, the aspirations acknowledge the power of high-quality data, that it should be used with confidence by decision-makers and partners for insights for responding quickly to threats, opportunities for prevention, and supports data-driven processes to keep the public safe. It lays out five key requirements:

- Tactical and operational decision-making,
- Performance reporting,
- Investigations and analysis,
- Forecasting and prediction,
- Strategic decision making,

which requires (a) easy, reliable, and fast access to data, (b) automated routine and repetitive activities, and (c) 'smart technology' to proactively find and surface information and insights. Data maturity[4] scores and targets are given (see Figure 1 in Appendix), and data principles that underpin success criteria are proposed (see Table 1 in the Appendix). Capabilities required (including data and ICT expertise), governance roles, and a roadmap is laid out over a 36-month period.

As part of the data strategy, semi-structured interviews with a broad range of over 140 police staff across the country were conducted. The findings from the interviews informed the current and future challenges that hinders NZP from achieving the aspirations set out in the NZP Data Strategy. Out of the interviews there were twelve themes that were extracted. Below is a summary of these themes:

1. **Querying, searching, and reporting:** Querying and searching is time consuming due to a lack of tools to search multiple data sources simultaneously. Reporting is standardised, but many staff found standard reports unhelpful, or what they needed from the reports is not available.

2. **Data capture and quality: Police collect** a wide range of information, but it is not well documented as to why we want it, what we intend to do with it, and whether it is collected at the appropriate time or what the best way to collect it is. There are technical causes and behavioural causes for the poor quality of data.

3. **Inconsistency and lack of standardisation:** This creates unnecessary work makes systems and processes complex and impedes data interoperability.

4. Data maturity is a measure of how advanced a business or agency's data ecosystem is, and is generally scored between 1 and 5. See https://www.data.govt.nz/leadership/data-maturity-assessment/ for more information.

4. **Information and data sharing and security**: The challenge of sharing data is a big concern, and barriers exist for sharing data internally and externally. Security is also a concern.

5. **Information provenance:** There are a few main systems of record for data, but information can come from multiple sources (e.g. offender profile in NIA). When the source is not known or not understood, this can lead to errors in reporting and interpretation.

6. **Unreportable data:** Some applications we use for capturing data and information predate modern reporting requirements. Some new applications were not designed with reporting in mind. It is labour intensive to extract data and gain insights from some applications.

7. **Duplicating databases:** The want for new tools can create new data sources, but these often just duplicate existing data, convoluting systems which makes it harder to find and use data.

8. **Descriptions of data (metadata):** Analysts require descriptions and definitions of the data that they access to use and interpret correctly, but there is little readily accessible documentation of the data.

9. **Systems and applications training:** There is a lack of training routinely available for most applications. Staff rely on colleagues to show them how to use most tools and applications.

10. **Old software versions:** Some of the specialist applications are older versions, which comes with technical and security risks.

11. **Subject matter experts (SMEs):** Many datasets have SMEs who know and understand how to use and interpret the data they hold and manage. However, it can be difficult to identify who the SMEs are and there is a risk that their knowledge will leave with them if it is not documented sufficiently.

12. **Integrity and Professionalism:** Analysts have little in the way of professional protection and can feel vulnerable to pressures to conform.

From the insights above, the Data Strategy acknowledges the following challenges:

- Data is not findable and accessible – unsure what data exists, where it lies, or what it means.
- Data is not interoperable – it is difficult to match data from different datasets.
- Data lacks trust – hard to trust the reporting due to differences between datasets.
- Data rich, but insight poor – hard to generate quality insights with current data.
- Cannot move forward with the data – cannot advance the use of automated systems and ‘smart’ data-driven technologies with poor data.

Data SaGA Interview and Workshop Findings

We conducted interviews and workshops with over 70 people. These included staff within NZP and people outside who worked for government agencies, independent researchers, and universities. Many staff inside NZP that participated were data practitioners and ICT experts, but also included those with experiences of front-line officers and 111 and 105 operators.

Interviews were semi-structured which allowed participants to speak on their roles and freely about the issues. Workshops were used to gauge participant insights into policing services, the data gaps that they know or believe exist, and how to overcome these data challenges.

Participants were also encouraged to share any other information they thought would be important to these issues. Many of the concerns raised here were similar to the concerns and comments raised in the Data Strategy interviews. We present the high-level findings below in no particular order.

The need for demographic data

Participants involved with data practice within NZP and projects regarding policing services almost all agree that demographic data is extremely important for Police to have, for insights, reporting, and measuring the effectiveness of Policing services across different groups and communities. In many cases, Police have obligations to capture better quality data for demographic variables such as gender, ethnicity, and disability data. Through the interviews, many participants bought up the fact that there were limitations with the current state of Police data pertaining to gender, ethnicity, and disabilities data.

Gender

Currently, according to the NRS, NIA allows the recording of gender as “Male”, “Female” or “Unknown”. The binary option is insufficient and does not comply with the standards set out by Tatauranga Aotearoa Statistics NZ (hereafter referred to Stats NZ)[5] or Te Kawa Mataaho Public Service Commission[6]. Without changes to these standards, this limitation in the recording of gender may inhibit the ability for Police to demonstrate appropriate and respectful engagement and meet the needs of gender diverse communities throughout NZ. Use of ‘gender’ is often confused with ‘sex’ and Police systems reflect a very binary and out-of-date approach which does not align with Stats NZ definitions of gender.

5. Data standard for gender, sex, and variations of sex characteristics | Stats NZ
6. Guidance-Workforce-Information-for-State-Sector-Agencies.pdf (publicservice.govt.nz)

Ethnicity

Similar to gender, there are challenges with the ethnicity data that Police records and that NIA holds. The ethnicity of a person is selected according to just one option they most feel strongly aligned to, with the options being "Asian", "European", "Indian", "Latin American", "Maaori (sic)", "Middle Eastern", "Native American", or "Other[7]". It was noted from several participants that ethnicity as captured by Police has evolved from 'race' as a concept and the current usage retains the original lens of a physical descriptor. Therefore, the list of values to select from does not reflect the current definition of ethnicity as determined Stats NZ standardised definition of ethnicity. Several participants were also concerned about the level of self-reporting of ethnicity, with estimates ranging between 2 - 5%.

Regarding the ethnicity categories, there are no set definitions for these within the NRS, so the way that people apply them may vary. This data is supposed to be collected in a way which the person self-identifies. However, there is no training for staff on how this is done. Once ethnicity is recorded once, it stays there. It can be changed but you may never be prompted to change it. The databases are quite old and have been around since the 1970's, so if someone was incorrectly identified 40 years ago it is still sitting there in the database.

7. "Other" has a free text field to record an ethnicity.

Disability

A stocktake of disability data within Police was undertaken in 2021. It was found that the overall level of data quality was poor. There are no standard recording processes, and nothing in the NRS as to how disability data should be collected or recorded. A report from the Police Disability Data Governance Group (PDDGG) in 2023 note that Police's definition of a person with disabilities comes from the United Nations Convention of the Rights of Persons with Disabilities,

"Persons with disabilities include those who have long-term physical, mental, intellectual, or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others."

The findings showed that for the most part, NZP has the legal authority, government mandate, and social licence to collect, retain, and share disability data for both operational and non-operational purposes. Participants speaking to disability data state that Police acknowledge NZP need to do more to engage with disabled communities but have little knowledge if they are doing so appropriately now and know they have done poorly in the past.

Improving disability data is very much needed to improve NZP's understanding and responsiveness to the disabled communities. Currently, there are initiatives to gain some data through NZP's Ō Whakaaro, Ā mātou mahi | Our Service, Your Say (OSYS) survey, and the NZ Crime and Victims

(NZCVS) survey run by the Ministry of Justice. The Disability Data and Roadmap is an initiative to address problems regarding disability data at the time of the interviews and laid out 18 action points to help NZP move towards better gathering and use of disability data.

Data Capture

Many participants made it clear that there were large gaps in the demographic data NZ Police hold. They were also clear that data collection is an issue that is challenging to solve. Most of the data that is captured and recorded come from frontline staff and communications operators.

Many participants agreed that staff could "tighten up" data collection practices. Training may be needed to help frontline staff collect data from the public in a better way, and to help record captured data within NIA and other databases. Human error in data entry is always an issue! Some participants felt that various aspects of the NRS are still too loose, and constant updating of the document is needed to ensure some data integrity. Updating data systems to ensure that there were appropriate options for demographic data was mentioned several times as a priority that could help data capture.

However, it was also made clear by participants that we should not overburden staff with capturing data that Police currently do not have. Frontline police attending an incident are often placed in highly stressful situations with members of the public who are stressed. Communications staff may answer a call where minutes or seconds may determine life or death. Collecting ethnicity, gender, or other data can get in the way of core services in some instances.

Some participants saw aspects of demographic data as "nice-to-have" for policy and deeper analysis, but from an operations perspective, is less important than the data required to attend an incident in a safe and timely manner. Police collect data as a function of their services, but their role in data capture should be limited to collection of core data.

Practical limitations also exist. We mentioned the limitations of collecting data for roadside stops. Data regarding victims can be a "huge challenge". An estimated 40% of victim's information cannot be collected at the time of an offence, mainly due to the complexity of the situation. Often, victims cannot be interviewed, such as when someone reports a crime happening to a victim – an offender may be caught and reported, but the victim has left or fled the situation and cannot be interviewed.

Many participants who were involved in the governance and use of the data acknowledged a fundamental issue, that there is a misalignment between data for operations, and data required for research insights, and policy. Data is collected by Police as a function of their service, such as occurrence or offence type, time, location etc. (it is possible to have over 50 fields, yet on average around 20 are filled and recorded), and data capture is designed for operations in mind. However, the data collected is not sufficient for policy and research. With the lack of demographic data, there are few chances to perform quantitative, empirical-based research on service delivery across different communities.

Data Quality, Reliability, and Validity

Past audits have shown that the data quality is poor, though improvements have been made over time. Participants noted that there are data assurance processes and that audits are performed often, but not over the entire country. Those involved in this area describe it as in a "fledgling state" and more is required to improve the quality, reliability, and validity of the data.

There are limitations of frontline staff collecting data on the ground. As stated above, tense situations can lead to information being hard to gather, and data that is captured may not be true.

Often data is hard to validate and requires a lot of time to check. Four staff work full time just to ensure hate crime data is reliable. It is uncertain how much data quality affects the quality of

Police official statistics at a national level, but the effect of poor-quality data will greatly affect the quality of statistics as we disaggregate down into smaller population groups.

Data Systems

Many participants commented on the inefficiencies of the current data systems and infrastructure. The databases currently in use are "old" and do not easily connect with each other. The databases have been in place since the 1970's and have been built upon over the years creating a siloed infrastructure, making data hard to find, access, not interoperable, and therefore not optimised for both operations and research uses. This creates an environment where internally, staff must extract data from multiple sources manually, which leads to datasets being emailed, carried on USB's etc., which is poor practice from a security perspective.

Participants commented that the system does need a complete overhaul to bring it into a more modern state but recognise the scale of this job is large and expensive. However, with the inefficiencies present in the current systems and the cost of maintenance, it is expensive to keep the current data system and infrastructure as is.

Data Sharing

With the gaps in demographic data and the limited ability for frontline staff to collect data that is required, many participants felt that the collection of data required may be best captured through other sources. Some mentioned that surveys are a good way to capture information, citing the NZCVS survey as an example of obtaining good information about victims. Those that have worked with the survey note that it is very hard to obtain information when disaggregating down into smaller demographic groups, even with the large sample size.

Many participants suggested that data gaps for individuals could be filled using data from other agencies, if they can be identified. For example, rather than Road Police collecting data, better and more accurate information may be obtained about an individual if we get information from Waka Kotahi (NZ Transport Agency), Te Tari Taiwhenua (Department of Internal Affairs), or Te Manatū Whakahiato Ora (Ministry of Social Development (MSD)). All these agencies collect data under far less stressful situations than Police, and MSD have far better standards when it comes to collecting and recording gender and ethnicity data. They also enable a person to self-identify their gender and encourage users of their services to actively search for their own ethnicity details, including iwi and hapū information, and provide information about how to go about doing so.

However, it was noted by some participants that data sharing comes with its own set of issues. Other agency's data systems and infrastructure are different to NZP thus the act of sharing data can be complicated. Security issues may arise if data is shared manually. For some services, there have been attempts at multi-agency approaches with data used to inform services.

In 2017, the Family Safety System prototype database was developed, for which the purpose was to combine data from multiple agencies to inform staff of family harm incidents. Unfortunately, there were large issues regarding the database, and its lack of synergy with other agency's data systems. Participants asked the question as to if we really want Police to be holding other agency's data at all. If you can build the infrastructure to share data efficiently,

Police have enough work holding and managing the data they have now. It may be far more efficient if the responsibility of filling data gaps for particular research purposes was given to the researchers to fill. In that case, Police would need an up-to-date data catalogue along with the NRS, so that researchers know what data is held by NZP, where it lies, and information regarding the collection processes.

There were suggestions to look at using the NZ Integrated Data Infrastructure (IDI) to source demographic data. The IDI currently has eight broad categories in the IDI:

- **Health data** – spans a wide range of datasets including cancer registrations, chronic conditions, B4 school checks, pharmaceuticals, mental health and addiction, laboratory claims, mortality, and more.
- **Education and training data** – includes education levels from early childhood education participation, through primary, secondary, tertiary, adult competency assessments, and industry training.
- **Benefits and social services data** – includes data on benefits, youth services, CYF[8], ACC[9] injury claims, student loans and allowances, and family start.
- **Justice data** – includes microdata from Ara Poutama Aotearoa (Department of Corrections), Ministry of Justice, and Police records.
- **People and communities' data** – includes data from Auckland City Mission Auckland Transport (driver's licences and motor vehicle registrations), as well as data from the following surveys: Immigration NZ's migrant surveys, Longitudinal Immigration Survey of New Zealand, general social survey, disability survey, and Te Kupenga.
- **Population data** – contains information on border movements, visa applications, departure, and arrival cards, as well as personal details such as births, deaths, marriages, and civil unions.
- **Income and work data** – includes microdata on tax and income, as well as survey data on income (from New Zealand income survey); household labour force survey (HLFS); survey of family, income, and employment (SoFIE); and the household economics survey (HES).
- **Housing data** – includes tenancy and social housing information.

NZP provide three sets of data in the IDI, Recorded crime: offenders, Recorded crime: victims, and NIA links. As of October 2023, both offender and victim datasets are up to date. However, NIA data in the IDI has not been updated since February 2018[10], though efforts are currently being made to update these. Some interviewees thought that the IDI could be the best way to fill demographic data gaps if identities can be matched to records which is the advantage of the IDI. Although the IDI may help fill data gaps, a lot of the administrative data collected by agencies are collected for their purposes.

Depending on questions researchers may want to ask, finding the right data retrospectively from sources of administrative data may not always yield the desired results (this comes back to the misalignment of operations and research). But if possible, using the IDI could fill demography data gaps and relieve some of the burden that those on-the-ground face when trying to collect that data.

8. CYF — Child, Youth, and Family
9. ACC — Accident Compensation Corporation
10. See https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/data-in-the-idi/.

Māori Data & Data Sovereignty

Several Māori staff participants raised the issue of the use of Māori data[11], and Māori data sovereignty (MDSov)[12] within the NZP's data ecosystem. The Data Strategy noted in its set of data principles (see Appendix 1) that:

The data we hold belongs to the people of New Zealand. We are authorised to use it to advance the safety and wellbeing of all New Zealanders. We treat the data we hold about Māori and ethnic groups with special consideration. We recognise its cultural importance. For Māori data, we work to adopt the principles of Te Mana Raraunga[13] and to support the goals of Ināia Tonu Nei[14].

Participants wondered how the principles underpinning MDSov were adopted into the data systems and data practices of NZP. For example, if an offender has a tā moko (tattoo) where iwi or hapū information is embedded, what are the considerations for holding that data? Only if it is required for evidence, or not at all? It was suggested many times that frameworks that establish the culturally appropriate use of Māori data in NZP data systems should be developed, similar to the Stats NZ Ngā Tikanga Paihere framework[15]. Some participants with expertise in Indigenous data sovereignty noted that, though the aspiration of building data systems where data is FAIR (Findable, Accessible, Interoperable, and optimised for Reuse), this has led to research practices that have disadvantaged already marginalised communities. From a broad Indigenous perspective, FAIR principles should be balanced with CARE principles[16] to ensure ethical use of data.

Other participants were concerned about historical injustices and systemic biases that Police have had towards Māori, and the effects of deficit-framing using crime and justice data impacting on Māori dignity, and how this results in the detriment of whakapapa. Several questions regarding these issues were raised by participants. Will more efficient data systems lead to more efficient perpetuation of systemic bias? Will sophisticated models of crime in time and space lead to more sophisticated surveillance of Māori communities? Can we ensure the benefits of better policing service result in good outcomes for Māori? Storytellers using data and statistics have often spoken on Māori and their disproportionate numbers in crime and prison, but often leave out important historical context of colonisation, social prejudice, and racism. If we can achieve accurate ethnicity data leading to models and statistics, who will be involved in communicating what the data says? Do, or will Māori have active participation in these processes? Acknowledgement of historical injustices and the role data has played in this, clarity of how NZP will use Māori data moving forward, and allowing Māori to be a part of the process in reclaiming their data rights may be a step towards a more positive relationship between NZP and Māori communities.

The Public Records Act 2005 and the Information and records management standard supports the rights of Māori under Te Tiriti o Waitangi/Treaty of Waitangi to access, use, and reuse information that is important to Māori. This may include enhancing metadata to make information easier to find by or for Māori or ensuring that information of importance to Māori (e.g., information about people, natural resources and land, or information required to support specific Te Tiriti commitments) is easy to access and use.

11. Māori data is defined as data that is about Māori culture (including art, language, history, etc.), Māori people, or environments that Māori have rights or interests in.
12. MDSov is the idea that Māori data should be subject to the laws and governance of Māori. See https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/5bda208b4ae237cd89ee16e9/1541021836126/TMR+Ma%CC%84ori+Data+Sovereignty+Principles+Oct+2018.pdf
13. Te Mana Raraunga is an advocacy group for Māori rights and interests in data. See https://www.temanararaunga.maori.nz/
14. See https://www.inaiatonunei.nz/about
15. For more on Ngā Tikanga Paihere, see https://data.govt.nz/toolkit/data-ethics/nga-tikanga-paihere/.
16. More on the CARE (Collective benefits, Authority to control, Responsibility, Ethics) principles can be found at GIDA website, https://www.gida-global.org/care

Recommendations

Based on the literature consulted and the findings of the interviews and workshops, we present a set of recommendations to help fix data gaps within the Police data ecosystems based on the participant interviews, workshops, and literature reviews of various documents such as the NRS and Data Strategy.

The first set of recommendations speak to improving data practices and systems within Police. The second set of recommendations are to help fill data gaps for quality research and evidence-based policy development. The final recommendation speaks to further data infrastructure developments outside of Police.

Improving Data Systems and Practices Within NZ Police

Recommendation 1

Development of a NZP Data Catalogue that sits alongside the National Recording Standards, to inform what data the Police currently have, where it can be found and accessed.

Given the issues surrounding findability of the data that came through in the Data Strategy findings and the Data SaGA interviews, a data catalogue outlining what data is held, where it lies, and details around the accessibility of the data, would be a valuable resource for all. Given the dynamic nature of data, data systems, artificial intelligence, and emerging technologies, the data catalogue would be a living document, and provide some context about the data itself — why it exists, and some of the limitations surrounding its collection and use. It should be a reference point for data definitions and standards (along with the NRS), what data should be captured, how it is recorded, managed, stored, and used. Although a complete overhaul and replacement of the current data systems with a modern data infrastructure is very much needed to help findability and accessibility of NZP data, a data catalogue would help with these issues in the short-term.

Recommendation 2

A formal set of guidelines and strategy for the appropriate governance and use of Māori data within the Police data ecosystem.

A framework for the appropriate governance and use of Māori data could be similar to Ngā Tikanga Paihere (data.govt.nz) for the context of NZP and its data ecosystem. Currently, under the principle of data sovereignty in the NZP Data Strategy, it is stated that they will adhere to the principles set out in Te Mana Raraunga[17] and supports the goals of Ināia Tonu Nei. However, details about what this means, and how it applies in practice is unclear. For example, how do we apply the principle of Rangatiratanga (Authority — Control, Self-Determination, Jurisdiction)?

The principles set out in Te Mana Raraunga are the cornerstone of Māori data sovereignty, but newer frameworks have been developed to guide governance and use of Māori data for specific situations. Since Māori data collected by NZP are generally stored in large databases, and use of

17. He Matapihi ki te Mana Raraunga – see https://researchcommons.waikato.ac.nz/handle/10289/11814.

the data is likely to be secondary use, Te Mana o te Raraunga framework may be a more applicable framework to follow. This work should involve NZP staff, including data practitioners and Māori staff within Police, and Māori data sovereignty advocates such as Te Kahui Raraunga or Te Mana Raraunga.

Recommendation 3

A larger "data workforce" for efficient and accurate data capture, data entry, and data quality measures.

All participants acknowledged the complexities of capturing the data that is required to better inform operations, and to gather research insights to inform strategy and policy. It is clear that those that are working to ensure data is of high quality, is reliable, and valid, require more support to achieve the aims and aspirations of the NZP Data Strategy. As we move forward into an (even more) data driven world, building an infrastructure that supports the aspirations of NZP requires investment — not just in systems, but in people.

A "data workforce" on the ground that supports frontline officers to capture data, enter data, and perform data quality measures and audits is required to ensure data quality is of high quality, is reliable and validated for use. This would also relieve some of the burden put onto front-line officers to ensure they can perform their core policing roles, and use the data effectively to support their work.

Recommendation 4

Building data systems to ensure data is findable, accessible, and interoperable to ensure optimal reuse of data.

Data systems that hold data that is not findable, accessible, and interoperable, cannot be optimised for reuse. For NZP to achieve its data aspirations, it needs to be supported with an infrastructure that ensures data can be reused well, along with platforms that are easy to use. The development of modern data infrastructures is a huge cost and is not something that can be developed overnight or reasonably in the short-term. This recommendation should be regarded as an aspiration to be achieved in the long-term.

Filling Data Gaps for Quality Research and Evidence-Based Policy Development

Recommendation 5

For cases where data is required that Police cannot collect or do not collect at all, it must be the responsibility of the researcher to fill gaps. Recommendation for greater use of statistical surveys to obtain insights that cannot be obtained with current data gaps.

Data gaps will almost always exist for researchers, especially when there is research being done on particular communities that are subsets of the

18. MVPFAFF+ - An acronym to describe Pasifika identities; Mahu (Hawai'i and Tahiti), Vaka salewa lewa (Fiji), Palopa (Papua New Guinea) Fa'afafine (Samoa) Akava'ine (Rarotonga), Fakaleiti (Tonga), Fakafifine (Niue). Other terms include Fakaleiti, Rae rae, and Fafafine.

groups that Police collect data on, such as those of African descent, hapū, LGBTQIA+/MVPFAFF+. In some of these cases it may be better for researchers to conduct a survey rather than relying on Police to collect it. There are several reasons this may be:

- Data collected by Police can be deficit framed. Collected from an independent (from Police) survey may be more positively framed, or framed from the community's perspective, leading to better quality insights.
- Surveys performed with sound statistical methodologies can extract excellent information with less data, decreasing costs.
- Reduces the load on frontline staff to collect the data.
- Will ease some of the burden on the data system overall. There may be data that Police researchers need, but we do not necessarily want Police to hold in their systems, due to storage costs, or the data is outside of core police business.

The data catalogue, suggested in Recommendation 1, would be an excellent resource to help any researcher understand available data and what data is needed to support and complete their research. The IDI could help researchers fill gaps if matching records to identity data is possible.

Further Data Infrastructure Development (Outside of Police)

Recommendation 6

Working with other agencies to build data infrastructure that makes data findable, accessible, enables efficient data linking and interoperability, to improve operations and can help researchers plug data gaps.

In many situations, data is required by frontline staff that Police do not collect as part of their core business, and thus, the data must come from different agencies. Initiatives to develop systems that incorporate Police data with other agencies' data have been tried but have not been successful. However, we encourage these initiatives of a multi-agency approach to improve both operations and research insights that come about from sharing data. Though sharing data creates many opportunities, it does come with some risk. Necessary precautions must be taken to ensure data remains private and secure.

Generally, the data infrastructure in NZ is inefficient, not just Police. The IDI was developed to increase the capacity for evidence-based research to improve social outcomes in poverty, health, education, crime, and overall wellbeing. As part of the IDI, NZP must maintain their role in the development and maintenance of the IDI, including regular updates on the NIA database within the IDI.

Appendix

Appendix

#	Principle	Description
1	**Strategic Asset**	We value our data as a strategic asset that supports and enables us to reach our corporate goals and deliver our legislative functions. It provides value to us, to our partners, to government, and to all New Zealanders.
2	**Data Sovereignty**	The data we hold belongs to the people of New Zealand. We are authorised to use it to advance the safety and wellbeing of all New Zealanders. We treat the data we hold about Māori and ethnic groups with special consideration. We recognise its cultural importance. For Māori data, we work to adopt the principles of Te Mana Raraunga and to support the goals of Ināia Tonu Nei.
3	**Open**	We make our data available by default unless it needs to be protected. We make it available freely and at no cost where possible to those who have an interest in or need for it. This applies internally and externally.
4	**Protected**	We protect all our data from unauthorised access. We are particularly careful with personally identifiable information (PII) and other types of controlled data to ensure we manage the risk to individuals from compromise of this information. Security of our data is integrated into its management throughout its life cycle.
5	**Accessible and Timely**	We make our data available at the earliest opportunity to internal and external users. We make available as much data as we can, in formats that are machine readable and easy to use.

#	Principle	Description
6	**Authoritative and trusted**	Data supports the purposes for which it was collected and is accurate, relevant, timely, and consistent. Wherever possible there is an identified authoritative single source (master), and reuse of the data is from that master.
7	**Comparable and Interoperable**	Our data is digital by default. It is captured and stored in ways that facilitate reuse and adheres to standards that ensure interoperability. Whenever possible it is available at its most granular. Where data needs to be protected through aggregation, it is aggregated at the lowest level possible.
8	**Defined and described**	Our data is fully described and defined with complete, up to date and accurate metadata. We use a common vocabulary and consistent definitions, which are readily available to all data users.
9	**Compliant**	Our data management practices ensure we are compliant with relevant legislation. We use Cloud storage when it is safe and compliant with relevant legislation.
10	**Well managed**	Data is continually managed throughout its life cycle, including catering for technological obsolescence and long-term preservation and access. We govern our data to ensure compliance, manage risk, and maintain consistency with good practice.

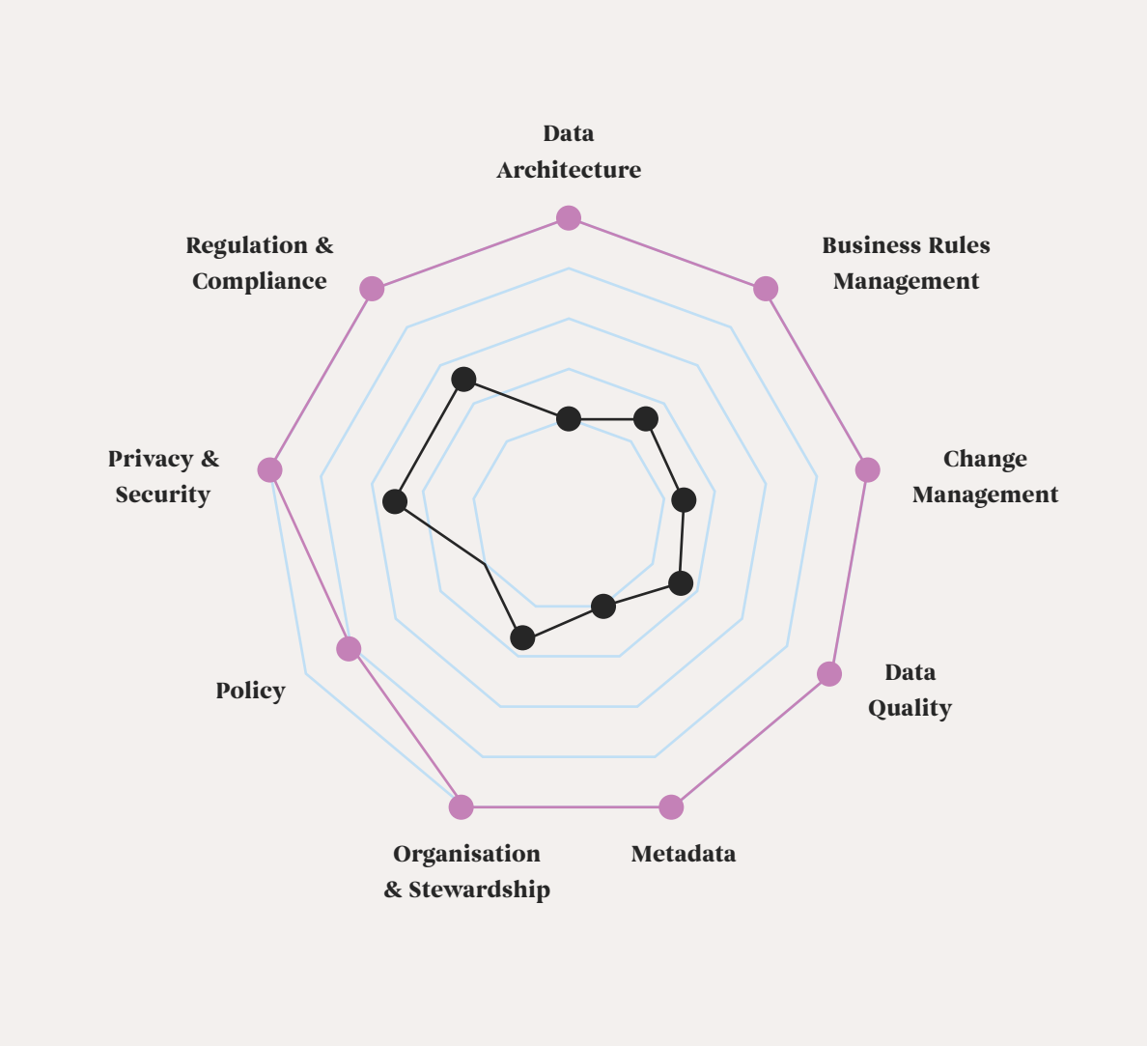


Figure 1: Data maturity targets for each facet of the data strategy (as at September 2021).

UPD

UNDERSTANDING POLICING DELIVERY