OPUS

Public Attitudes to Data Integration

Report prepared for Statistics New Zealand

OPUS

Public Attitudes to Data Integration

Report prepared for Statistics New Zealand

Opus International Consultants Ltd
Opus Research
33 The Esplanade, Petone
PO Box 30 845, Lower Hutt 5040
New Zealand

Telephone: +64 4 587 0600
Facsimile: +64 4 587 0604

Anna Davison, Jean Beetham, Jared Thomas, Abigail Harding, Vivienne Ivory, and Chris Bowie

Date: 3 July 2015
Reference: 5-27813.00
Status: Final

Approved for Release By

JThomas

Dr Jared Thomas
Research Manager, Behavioural Sciences

Executive Summary

The purpose of this work is to better understand public attitudes to data integration and the willingness of the New Zealand public to share their information directly or indirectly. More specifically, to understand how individuals from different demographic groups might perceive data re-use and integration, reveal any differences in perception of operational and statistical uses of data, and compare acceptability thresholds around this data use by the public. A further purpose is to examine the situations in which improved information might influence acceptability thresholds around data use, and to identify any barriers to confidence in Statistics NZ as a custodian of integrated data. By gauging the level and extent of public concerns, the reasons for those concerns, and the important distinctions between them, public sector agencies will be better positioned to anticipate public requirements and respond appropriately.

A multi-method, qualitative approach was chosen to investigate these complex concepts across a diverse range of groups. The in-depth qualitative approach also ensured participants understood the processes of data integration, so that they could engage thoughtfully and in greater depth with this topic. The qualitative research comprised of three stages, including: 1) in-depth narrative interviews in Wellington and Christchurch; 2) workshops across three urban centres (Wellington, Gisborne and Invercargill); and finally 3) an on-line discussion with different experts. Recruitment for the interview and workshop stages of the research actively targeted social groups of interest, including Māori, Pasifika, retirees, the unemployed, and the self-employed, but also included New Zealand residents who did not belong to these groups.

This research found that, fundamentally, data integration acceptability appears to be largely influenced by individuals' own personal experiences with, and trust in, operational data sharing, the formation and use of statistics, as well as government agencies. Based on the research findings it is apparent that general data integration acceptability thresholds do exist. Participants appeared to gauge acceptability primarily based on the need for the information, how the data would be used and by whom. They were also interested in the value of the data integration, whether the benefits would be greater than the costs and risks, and how any potential risks might be mitigated.

Data integration tended to be seen as relatively more acceptable if there was a clear and appropriate need for it, if it was in the public interest, and it produced positive outcomes that had individual or public benefit. Participants' emphasised that organisations involved would need to be highly trustworthy, professional, and competent. Strict procedures and protocols would need to exist around access and use, and databases would need to be completely secure and confidential. Statistical data integration was more acceptable if the data is completely de-personalised and anonymous. People tended to feel that data integration is more acceptable if the public is informed about what is happening and why. If personal, sensitive, or complex data is going to be integrated some participants felt that people should have the right to give or withdraw their informed consent.

Data integration tended to be seen as relatively unacceptable if there was no demonstrable need or purpose. People also tended to feel that integration would be unacceptable if it could be misused, or resulted in harmful or unfair outcomes. Examples of such outcomes include: if integrated data was used for direct commercial gain; to take advantage of vulnerable people; or for profiling, stereotyping, or disadvantaging certain groups or types of people, or people from particular places. People felt uncomfortable if poor quality data was used in a misrepresentative way, and some were unhappy about data being used for spying and surveillance.

If data was integrated in an acceptable way then most participants felt that this would be valuable and beneficial. People felt that integrated data systems could be more reliable, current and accurate than those currently used, and could result in more informed, fair, efficient and effective decision-making and service provision.

Most of the research participants had a high level of trust in Statistics New Zealand as a professional, competent and trustworthy custodian of data. However, some of the research participants had low levels of trust in Statistics New Zealand. Those with low levels of trust tended to feel this way if their experience with statistics was negative, such as if they felt that they personally, or where they live, had been stereotyped, stigmatised, or disadvantaged by generalisations based on statistics. Some also felt that statistics, in general, were not a useful or meaningful way to assess how well people are doing, or how happy they are, in a broader sense.

In order to retain and gain the public's trust, Statistics New Zealand should take every care to actively demonstrate that they meet the acceptability criteria expected by the wider public if, and when, they conduct data integration. A summary of the core concepts is simply outlined as follows:

Participants appeared to gauge acceptability primarily based on:

- *Why* the data needed to be integrated,
- *How* any personal or sensitive information would be protected,
- *What* the data would be used for, and
- *Who* would have access to and use the data
- Whether the benefits and value would outweigh the costs and risks

Data integration tended to be seen as relatively <u>LESS</u> acceptable if:

- There was no demonstrable need or purpose: the general sharing of all information was typically seen as neither necessary nor acceptable
- If it could be misused, or resulted in harmful or unfair outcomes, these include:
  - direct commercial gain
  - taking advantage of vulnerable people
  - profiling, stereotyping, or disadvantaging certain groups or types of people, or people from particular places
- Subjective, biased or limited data was used in a misrepresentative way
- It was used for spying and surveillance

Data integration tended to be seen as relatively <u>MORE</u> acceptable if:

- The need for and purpose of integration was specific and transparent
- It aligned with the greater public benefit and interests, and improved social outcomes
- The organisations involved were seen as highly trustworthy, professional, and competent
- If the data, metrics, integration process, and resulting uses were: fair, appropriate, meaningful, accurate, representative, and of high quality and integrity
- Strict procedures and protocols exist around access and use
- Databases are completely secure and confidential
- Data is completely de-personalised and anonymous for statistical uses
- The public is informed about what is happening and why
- People are freely able to give and withdraw consent for their data being used, particularly if the data is personal or sensitive

Summary infographic of core factors that influence acceptability

Less acceptable	↔	*More acceptable*
Ambiguous & general	Purpose	Transparent & specific need
Insecure	Security	Secure
Open access	Access	Strict restrictions, procedures & protocols
Public/ personalised	Privacy	Confidential/ anonymous
Private	Interests	Public
Subjective & biased	Data & metrics	Representative & meaningful
Unfair & harmful	Outcomes	Fair & beneficial

Perceived benefits of data integration

Participants felt that acceptable data integration could result in:

- Better quality data and therefore, better quality information:
    - Longitudinal data would be more readily available
    - Data would likely be more detailed and diverse
    - More accurate and objective data
    - Data that may be more current
    - Data that could be updated more regularly

- Operational data integration could result in considerably more efficient and effective public services
- Better information generated from statistical data integration may be used by researchers as well as government agencies to help with planning policies and services in a more effective way, which may also make them more efficient
- Better information may reveal more about peoples' needs and vulnerable populations which can help ensure these needs are able to be met

Recommendations

Statistics NZ could consider the provision of positive messaging around data integration to ensure they retain and gain the trust of the New Zealand public, including:

- Messages of reassurance around continual improvement in privacy for operational data and confidentiality around statistical data. Even basic information reassuring people that the information met a clear and appropriate need, in particular around a public good as opposed to a commercial need.
- Demonstrating that the benefits would be greater than the costs and risks, and how risks could be mitigated (e.g. such as misrepresentative use of data, or use of data for surveillance).
- Where the data is personal, sensitive or complex in nature demonstrate how people could be offered the right to opt out.
- Examples highlighting the existing public benefits of data integration, including real examples of how the data has been successfully used and by whom, to improve levels of service for individuals or communities.
- Examples highlighting the potential public benefits if data is better integrated, including the opportunities to make more informed, fair, efficient and effective decisions to improve quality of life for New Zealanders. For example, the ability to use wider or less traditional metrics when determining how well people are doing or how happy they are.

Suggestions for further research and actions

- Quantitative data could be collected to better gauge the strength and distribution of issues uncovered by the qualitative findings.
- Investigating ways of effectively providing public reassurance and allaying people's concerns around data integration.
- Developing a publication that clearly demonstrates how those acceptability criteria identified as most important are addressed by Statistics NZ.
- Developing a publication that clearly illustrates the different applications, uses and benefits of data integration.

Acknowledgements

The authors would like to acknowledge the inputs of the team at Statistics NZ, especially Che Tibby, Deb Potter, Lyn Kaye, Adam Brown, and Michael Holland. We would also like to thank the research participants for generously giving their time, energy, and thoughtful perspectives.

Contents

Executive Summary i

1 Introduction 1
1.1 Research Objectives 2
1.2 Process 2

2 Results: Narrative Interviews 3
2.1 Participant characteristics 3
2.2 Responsiveness to survey instrument 3
2.3 Understanding (knowledge about data integration) 5
2.4 Attitudes towards data integration 9
2.5 Acceptability 14
2.6 A highlight example of how acceptability can shift 18
2.7 Statistics New Zealand as custodian 18
2.8 Operational vs Statistical Use 19
2.9 Summary of common themes and differences across groups. 19
2.10 Key differences between groups 20

3 Results: Workshops 21
3.1 Participant characteristics 21
3.2 Responsiveness to the workshop instrument 21
3.3 Findings 21
3.4 Workshop summary 29

4 Results: Online Consultation 30
4.1 Participant characteristics 30
4.2 Participant response to the online consultation 31
4.3 Findings 32
4.4 Online consultation summary 37

5 Discussion 39
5.1 Acceptability themes and thresholds 39
5.2 Acceptability shifts 40
5.3 Factors that influence acceptability 41
5.4 The influence of research techniques on findings 43
5.5 Trust in Statistics New Zealand as a custodian of data 45
5.6 Data integration and the Census 45

6 Conclusions and recommendations 46
6.1 Suggestions for further research and actions 47

7 References 48

Appendix I: Research Methods 49
In-depth Narrative Interviews 49
Workshops 51
Online Consultation with Experts 55

Appendix II: Scenarios for Narrative Interviews 58

Appendix III: Personalised Data Blocks 61

Appendix IV: Scenarios for Workshops 62

Appendix V: Online Consultation Outputs 65
How to interpret the results tables 65
Scenario 1 65
Scenario 2 67
Scenario 3 69

1 Introduction

A key element of the Statistics 2020 Te Kāpehu Whetū (Stats 2020) strategy is to make greater use of administrative data. Statistics New Zealand's future data collection activities will take an 'administrative data first' approach, using data from a range of sources to extend and enhance the range of statistics provided. Knowing what the public think about Statistics New Zealand's data re-use and integration activities is important for the organisation to understand.

A variety of studies have been conducted that examine the integration of administrative data for operational use (Sherriff 2013 provides a review of relevant studies). Statistics New Zealand is interested in knowing the extent to which the findings of these studies also apply to statistical uses.

Statistics New Zealand does not yet know much about the willingness of the public to share their information directly or indirectly. What is known is that public perceptions have the potential to influence respondent behaviour. An important element in people's willingness to provide true and accurate information is that they trust the organisations collecting it, feel confident that it will be treated responsibly and that it will be used for legitimate purposes.

As Statistics New Zealand's efforts to integrate data accelerate, knowledge about public perceptions will better inform Statistics New Zealand's communications and understanding of public views. The visibility of current data linking activities is relatively low, but future plans are likely to attract greater public attention (e.g. Population Census). To prepare for that, it will be necessary to understand public perceptions, and importantly the distinctions there may be between different groups of the population and their willingness to share information for different types of activities.

In particular, Statistics New Zealand wish to understand:
- How individuals from different demographic groups of interest understand data integration, and how they react to newly-acquired knowledge of data re-use and integration;
- How people respond or react during discussion and exploration of topics in data re-use and integration; and,
- People's acceptability thresholds for data integration.

By gauging the level and extent of public concerns, the reasons for those concerns, and the important distinctions between them, Statistics New Zealand will be better positioned to anticipate requirements and respond appropriately. This will also help to inform communications that ensure continuity of public trust in the agency as a custodian of data, address unwarranted concerns, and raise awareness of the benefits data integration offers the New Zealand public.

Prior to the commencement of this study, Statistics New Zealand engaged a consultant to understand the agency's research needs and identify the general research approach. A three-stage qualitative research method was specified, including thirty in-depth interviews across two urban centres, a period of online consultation with subject matter experts, and twenty action research walkthroughs across multiple urban centres. Each stage of the research was intended to inform subsequent phases. It is recognised by Statistics New Zealand that qualitative research has the potential to provide a richer understanding of the topic and the range of issues that influence people's attitudes. Where quantitative research can be described as a long shallow dive into data that is limited in the amount of detail it provides; quantitative research can be described as a short deep dive into the topic, providing a rich and detailed understanding of the issues.

1.1 Research Objectives

The objectives of this research are to:

1. Gain insight into understanding of and attitudes to data integration generally;
2. Reveal any differences in understanding of and attitudes to operational and statistical uses of data;
3. Compare acceptability thresholds for operational and statistical uses of data;
4. Explore the potential for information to influence acceptability thresholds; and
5. Explore any barriers to confidence in Statistics New Zealand as a custodian of integrated data.

1.2 Process

The research comprised of three stages:

1. Thirty in-depth narrative interviews in Wellington and Christchurch;
2. Workshops with 30 participants across three urban centres (Wellington, Gisborne and Invercargill); and,
3. An on-line discussion to which 43 people were invited to participate.

Each of the three research stages involved scenarios to facilitate meaningful discussion. The team at Statistics New Zealand developed the scenarios and in some instances, these were refined by the researchers. For more on the methodology see Appendix I. The social groups of interest to Statistics New Zealand include: Māori, Pasifika, retirees, the unemployed, and the self-employed (Lips, Eppel, Cunningham & Hopkins-Burns, 2010). Recruitment for the interview and workshop stages of the research actively targeted these groups, but also included New Zealand residents who did not belong to these groups.

2 Results: Narrative Interviews

Thirty in-depth narrative interviews were undertaken in Wellington and Christchurch. The narrative format of the interview schedule provided a “safe” space for the interviewee to share any experiences they may personally have had with data integration and reveal their current understandings of and attitudes towards data integration. An intervention activity then focused on gaining more in-depth and quantifiable information about the acceptability thresholds that participants hold for a variety of hypothetical data integration scenarios. It also provided an opportunity to identify what, if anything, could be done to mitigate participant’s most serious concerns about the use of their data. See Appendix I for a more detailed methodology.

2.1 Participant characteristics

A total of 29 participants were recruited across the 5 target groups. Around two thirds of the participants were women and a third men. The age of the participants was spread between the ranges 25 to 65+. There was an even spread of participants across the different employment status.

Table 3: Narrative Interview Participant Characteristics

Target Group		n
Employment	Unemployed	6
	Self-employed	4
	Retired	5
	Other (e.g. employed, care-giver)	14
Ethnicity	Māori	6
	Pacifika	6
	Non-Māori, non-Pacifika	17
Age	65+	5
	25-64	24
Gender	Male	10
	Female	19

2.2 Responsiveness to survey instrument

The narrative interview schedule was an effective tool for encouraging discussion around peoples understanding and personal experience of data integration. It facilitated deep exploration of participant’s thoughts around acceptability in relation to various scenarios where the potential for both positive and negative outcomes was deliberately introduced into the discussion.

The ice-breaker question around form filling was extremely successful in encouraging the participants to engage in the topic area. It provided the opportunity to enter a potentially unfamiliar topic area with an example from their own personal knowledge and experience.

Several participants became immediately more openly relaxed and less apprehensive about their involvement in the interview having completed this exercise and having gained an appreciation of the research question and its implications for every-day life. As a result they were subsequently more confident in sharing their experiences and discussing the pros and cons of data integration as they had experienced them first hand. They were also comfortable with the idea that in any given circumstance there were often both benefits and potential harmful outcomes for either individuals or wider society, and that the absence of a 'right answer' gave them the freedom to explore the issues from a variety of perspectives without worry of being judged.

Participants were invited to relate a story around data integration that they or someone they knew had experienced, which was typically operational data integration. Their knowledge and understanding around this, and the purpose of the data integration, was then explored in greater depth. During the next stage of the interview process it was a helpful technique to adapt recently cited operational example and re-frame them as a statistical example in order to explore attitudes towards the use of data integration for this purpose. The participants had provided the original example and so it was already familiar to them, and they were able to more easily consider the different, but still relevant to them, perspective where their data, de-personalised and pooled with others could be used for a wider societal or group benefit.

One participant had, for example, related an experience around providing information for a working visa application. This context was then re-framed to enquire how the participant would feel, if, for example, their data had been pooled, in an anonymous format, with other visa applicants details in order to provide information to the government about the characteristics, skills, experience and education of people seeking working visas over a certain time period. Another participant related a story about how health data relating to her child had failed to be shared effectively. Her experience was then re-framed and the participant asked how she would feel if her child's health data was pooled, in an anonymous format, with information about other children with the same condition, in order for the Ministry of Health to assess the need to review screening protocols, on-going monitoring and treatment for children across New Zealand.

The scenarios generated a large amount of interesting discussion around the possible advantages and disadvantages of data integration, whether for individuals, the wider public or the particular organisations involved. Frequently participants were able to see that while there might be a benefit for one party, there could be a disadvantage for others.

Using the 'usefulness / acceptability' chart, made it possible to explore their acceptability criteria and encourage them to identify their boundaries and bottom line for these two elements. While some interventions were clear cut, others were much less so, sometimes depending on whether the participant had personal experience around the particular scenario that contributed to their perception of the issues.

The participants genuinely seemed to enjoy the interview process, many became enthusiastically engaged in the discussion and expressed interest in finding out about our high level findings in due course.

> *Please pass onto the interviewers that my members appreciated their involvement and they found their treatment was kindly and tactful.*
>
> \- *Chairman, SeniorNet Canterbury*

2.3 Understanding (knowledge about data integration)

Theme: Expectations about data integration
During the narrative interviews participants talk about methods of linking datasets revealed an expectation that integration between different government datasets already happens. How they talked about data integration suggests an overall level of comfort with this type of sharing information. The level of knowledge about and interest in specific data integration methods varied between participants, for example retirees and self-employed people tended to have a greater appreciation of the bigger picture and to have already spent time considering the issues compared to other groups who appeared to have more limited consideration of the topic area prior to participating in the interview. However, even those who did not appear to have previously given much thought to the issue of data integration prior to the interview were able to describe basic linkage methods and showed understanding of the possible processes and outcomes.

Knowledge of data linkage methods
There was a general expectation that individual information provided to Government departments is shared with other Government departments and that this is automatically conducted electronically behind the scenes.

Participants suggested that data is linked together by some form of identifier that is unique to a person, for example name, date of birth, or other personal details, “a barcode on your forehead” and that personal data is widely available across all government departments.

The most frequently cited linkages during the narrative interviews were between: -

- Inland Revenue Department (IRD)– Work and Income NZ (WINZ) – employers - ACC
- GP – hospital departments – health care providers – ACC – Ministry of Health - WINZ
- Police – Ministry of Justice – overseas and NZ immigration - spy agencies – Department of Internal Affairs
- Schools – Ministry of Education (MoE) – other agencies e.g. Child Youth and Family Services (CYFS) – parents – police

Knowledge of how confidentiality is maintained
Confidentiality was extremely important to participants who discussed at a high level two ways in which confidentiality needed to be maintained. Firstly, there was a general level of comfort with pooling data for statistical purposes where confidentiality was maintained by having the personal details removed. There was more concern about personal data being kept in records where there was felt to be a greater need for confidentiality, and control over data sharing.

The sharing of individual information by government agencies with private organisations or third parties was considered by many to be unacceptable unless the individual had given express permission for their information to be shared.

> *“Why would these other people want this information? Why would they need it if they are not part of the Ministry (of Education), or part of the Government”*

Some participants felt that their permission should be sought prior to sharing between different government departments, for example between health and education, or different health care providers, and that they should be able to decide what information was shared, for example, no sharing of historic data if it was felt to be unnecessary.

A few participants were concerned about what would happen to personal information including contact details, provided by them to an organisation who subsequently went out of business, was privatised or merged with a new organisation who might have different values or intentions for use of their data.

> *"If information is transferred through change in Government, information given to one organisation that is trusted could end up in the hands of another organisation who might not be trusted or have the same interests of the people as when the information was given"*

Others were concerned about the implications of data integration and confidentiality in small communities where everyone knows each other, and where people have access to information about other community residents through their work. They were particularly concerned about data being integrated from a number of different sources and then being accessed by someone for a different purpose who might know the individual concerned and have access to all their personal information across all government departments.

One participant had experienced her personal information becoming public following a court case that was reported widely in the national media and then disseminated further via social media, including photographs of her taken out of context. The lack of confidentiality around the reporting of her case had a huge impact on her personal and professional life and implications for her safety and reputation.

> *"Wellington is a small place and people talk. I have this reputation now in Wellington".*

Confidentiality in relation to the security of online data was identified as a concern for a number of people and a potential consequence of online data sharing. Those who had direct experience of hacking, for example, into their bank account, reported having high levels of trust in their bank to provide sufficient security and surveillance to detect and mitigate the impact of this for them.

Many recognised the need for individuals to take their own precautions when undertaking online banking transactions, for example when deciding which online shopping sites they used, and being careful to restrict who they provide banking information to in order to minimise the risk. Others simply stated that they would never use online banking services because of the perceived risks.

In a 2013 Office of the Australian information Commission report (cited by Sherriff 2013), 48% of people surveyed identified online services and social media as a major the threat to the security and privacy of their personal information.

Theme: Understanding the purposes of data integration

The purpose of data integration was clearly important to participants, suggesting they understand why data integration is being used. The majority of participants felt that most of the time data integration is used for positive outcomes and to benefit people, but they also voiced concerns around the potential for misuse to occur, albeit infrequently. This came back again to the primary themes of who was using the data and for what purpose, Government organisations being generally trusted to use the data for good reason, while the use of data by private organisations was less trusted and generally less acceptable.

Understanding of how data integration can be used to generate new information

Many participants had not considered how data integration could be used to generate new information prior to the narrative interview. The knowledge section of the interview helped the participants reflect on the opportunities presented by data integration for their personal benefit and to benefit the wider community.

> *"When I was diagnosed with diabetes, I was invited straight away to go to the hospital for an eye-test. You feel supported. …integrating data with the hospitals, why not do it, it should be done as a matter of course, come and have your eyes checked!"*

Thoughts around data being shared and then used to provide information about something other than the original purpose varied depending on what the data was being used for, who was using it, and whether the outcome would be beneficial or harmful for the individual or community at large. If the purpose of using the data was for the public good, for example to provide improved public services and support, or to assist with decision-making, or advise on a change of entitlement then it was considered acceptable for the data to be used in this way.

One example related to an online grocery shopping transaction, where participants were comfortable that data about their purchase history could also, and was probably being used for other reasons such as to provide alerts about special offers of relevance to them, to identify 'favourites' that would help personalise their next online shopping experience, as well as to monitor levels of stock and identify favourite brands that would assist the organisation with their marketing and decision-making.

> *"The supermarkets do it all the time to give you discounts"*

On the other hand the use of historic data taken out of context to inform a current situation was not felt to be acceptable, because it was felt that this might not take into account any changes in circumstances since the original data was collected, and could cause unnecessary distress or negatively impact on the individual concerned.

> *"My neighbour has breathing difficulties, emphysema through smoking, she was a smoker. They allocate resources now for her condition now, but it dates back to a time when smoking was acceptable, the health concerns were not known then. It shows a lack of fairness"*

Some participants talked about commercial organisations sharing or selling information to other organisations. For them this was completely unacceptable if used purely for commercial gain.

> *"I had to fill in a form for the NZ Post, they asked me for information to update my record, all of a sudden all these advertisements were coming to my email, then I saw on t.v that NZ Post were selling information to these companies"*

> *"I would object to an insurance company selling my information. I am dealing with them, the deal is between me and them, I don't want someone else coming in"*

Understanding of the personal and public benefits of data integration

During the narrative interviews the participants talked about a number of actual and potential benefits of data integration for both themselves and the wider public, suggesting that they have good knowledge about the purpose of data integration. These included; time saving, continuity of care, streamlining of administrative processes, detecting fraud, and notification of entitlements and services. These are discussed further in the 'Attitudes' section.

Understanding of the benefits to organisations of data integration

People recognised multiple benefits to organisations of data integration and were able to talk about a range of benefits that they considered both useful and acceptable including;

For organisations to generally;

- have more information about people
- better understand trends over time by grouping statistics about people
- monitor populations and people over time
- maintain up to date and accurate information and records
- compare levels of service provision provided between different organisations
- prevent duplication / repetition
- prevent crime
- create a knowledge base (e.g. for Ministry of Education around achieving national standards)
- create a bigger picture using data from a variety of sources

And more specifically to

- plan services
- inform decision-making (e.g. around funding, housing, benefits)
- provide information (e.g. entitlement to support services and benefits)
- identify need for services
- provide resources and services
- promote continuity (e.g. of care in health service provision)
- streamline administrative processes
- share information about job opportunities
- assist the immigration process
- prevent and detect fraud (e.g. making sure the right people receive their entitlements)
- ensure individuals meet their obligations (e.g. around tax)

> *"When I filled out the form for my gun license…I would have thought it would go on record and be shared to check who I am, to check that my ID checks out against my vehicle license, and to check with Ministry of Justice about my past history. The fact that they ask for the information makes me think that they then check out that it's all matching"*

Selling of information was also raised by some participants as a potential way for their information to be used without their consent and was highly frowned upon by many as not being an acceptable method of data sharing, but some considered it almost inevitable.

Participants recognised that their data could also be used by private organisations to

- inform marketing initiatives
- provide personalised shopping experiences
- provide landlords with tenancy history
- target particular individuals or groups for marketing
- exploit vulnerable individuals or groups

> *"Most information has the potential for a positive and negative outcome for different people. It depends who makes the decisions and who has the information"*

Understanding of the use of data integration for surveillance

Surveillance was mentioned during the narrative interviews by a small number of participants who suggested that there were a few organisations including the police, government and spy agencies who were able to monitor individuals by tapping their telephone calls, using the GPS data on their cell phone to locate them, and monitoring of their banking transactions for the purpose of surveillance of individuals of interest or specific groups of people.

Two participants thought they had been the subject of spying, both having concerns about national and also international data integration between the police and security services in different countries to monitor their activities.

> *"The police share information, they spy on you, they bug you, they share it with whoever they need to share it with"*

Those with no direct personal experience of surveillance felt it was useful and justified to use data integration for surveillance as long as a reasonable risk had been identified and that the person under scrutiny posed a real risk to society, when surveillance was considered to be for the greater good.

2.4 Attitudes towards data integration

Theme: Weighing up personal impacts of data integration

The positive impacts on individuals came through strongly in the stories narrated by participants during the interviews. The participants were able to talk about a number of actual and potential benefits of data integration as either experienced themselves, or in the experiences of others they knew of, that would have made a difference to the situations they related.

Impact of data integration on ease of activity

Effective data integration across agencies could make life easier for individuals. Participants gave many personal examples, including,

- continuity of care through health information being available to all providers
- co-ordinated care through health information being available to all providers
- informed decision-making around for example, where to live
- being reminded to take action for example, in relation to applying for tax refunds or completing tax returns, updating information affecting entitlements, attending healthcare check-ups, or completing a child's immunisation programme.

> *"the first time I did it (claim for tax refund on donations) I had to give them lots of information, but this time the IRD sent me the form, so that's the benefit, now they automatically send out the form"*

One participant has a daughter with high needs who receives input from a number of different health care providers. She talked about how data integration of her daughter's health records and therefore sharing of information between the various specialists allows better care to be received from the different health care providers, as they can co-ordinate joint appointments with more than one specialist at same location, with thoughtful and appropriate timing and scheduling of treatments or operations based on the child's other health needs. The participant appreciated the reduced burden on her as effective data integration of all her daughter's health information ensured all the different providers were kept up to date and informed.

Impact of data integration on saving time and resources

Data integration was also regarded positively because it saved time and resources for individuals, with stories recounting advantages such as;

- avoids the need for repetition, or duplicating information provision
- provides alerts to new job opportunities automatically
- provides all the available information, so there is no need to search for historic data
- provides notice of online shopping favourites, and specials relevant to the individual
- new information is automatically updated across other relevant departments
- administrative processes are streamlined

One participant gave the example where following the death of her estranged husband, integration of his death certificate, ACC record, kiwi saver and next of kin information enabled the holders of his kiwi saver account to locate his former wife and advise her of entitlements for the children from their father's kiwi saver scheme.

Examples were also provided of the consequences of a failure to integrate data that further shaped participants views that data integration could provide personal benefits. Frustration was expressed by participants about the individual wasted time and effort repeating the information giving processes, with the potential for information to be lost or overlooked.

> *"Our GE card was out of date, we hadn't used it for 10 years. We went to buy something... we had purchased from them (the store) before, and I was on the computer system, but I had to fill in 4 forms again, even though all the data was there. Data could have been integrated to save time and be more effective, they had all the information there already"*

One participant also talked about that while data integration worked so well for her daughter's healthcare provision, when it came to receipt of her daughter's entitlement to disability benefit, which was based on her having her life-long condition, she was still required to repeat the process of applying for her benefit, which included obtaining reports from her GP and other health professionals every 3 years, despite all the information being available and no change going to occur to her underlying condition.

Emotional Impacts of data integration

While the ease and convenience impacts of data integration provided positive, personal benefits, the emotional impacts tended to be negative. A small number of participants specifically talked about the relief of having the burden of responsibility taken away from them by automated administrative processes but most reported the negative outcomes of data integration they had experienced.

The emotional impacts of negative outcomes of data integration included:

- distress from damage to reputation through the use of historic data or data taken out of context

> *"After the court case...they hacked into my (facebook) account and took an unrelated photo of me at a party, someone at work saw it, it was in the news, it was shared to Auckland and my brother saw it in London. People at my partners work saw it and he wanted to separate from me. I was very emotional, distressed and suicidal"*

- anxiety due to personal information being shared publicly
- fear of information getting into 'the wrong hands
- distress caused if inaccurate information is shared
- feelings of unhappiness if residents in a community feel they have been labelled
- frustration if data integration leads to a loan application being declined
- anxiety due to uncertainty around change in entitlements
- emotional distress through having to prove a partners death to a utility company

  "Her husband was the bill payer, they were piling obstacles in her way. Did she have to produce a death certificate in order for them to fix her hot water? I was furious."

- guilt, anger, frustration if information is not shared in a timely manner with consequences for things such as the health or education of a child

  "I assumed my son had behavioural problems from when he was at kindergarten, but the school nurse picked up that he had difficulty with his ear, my GP got information from the midwife clinic from the day after he was born. They had noticed that there was disruption in his ear but it wasn't shared at the time. Now he has a hearing aid. I was upset at the start, he could have had an operation, but the data got stuck in transit, otherwise we could have done something about it before."

- distress through insensitive enquiries around entitlements following a death in the family
- annoyance at being made to feel that they had neglected to do something for their child
- fear and uncertainty if information has been shared when the individual doesn't know what information is being shared or who it has been shared with
- unfairness if an individual does not have access to the information held about themselves
- suspicion about what information is being integrated and the purpose of integrating
- uncertainty around the likely impact of the data integration for the individual concerned.

Impact of data integration on safety
Participants were able to talk about a number of practical and moral issues around the ability for data integration to impact on safety of the individual, with the potential for both positive and harmful outcomes. Many participants considered that the safety of the community as a whole should take priority over the needs of a single individual.

- positive impact of data integration between Ministry of Health and Ministry of Education ensured that a school received the necessary resources to make the school grounds and buildings secure and safe for a child with special needs starting school

- implications of data integration in presenting a moral obligation for those in receipt of the information to act on any safety concerns identified e.g. between school, health, child youth and family services (CYFS) or police, for the long term benefit of the child concerned even if it heightens the safety issues in the short term – those in receipt of the information having a duty of care to act on it.

- the potential for data integration to increase safety for children if school, police, CYFS, WINZ and health providers are all sharing information about a child, thus enabling the school to support the child and family for a positive outcome and put safety nets in place to ensure the safety of other children and families in the school community

- Implications of data integration and sharing for the safety of certain groups within the population / community in relation to their ethnicity, vulnerability, where individuals or groups could be targeted

"She might get robbed if someone knows she is on her own"

Impact of data integration on data security

Through the narrative interview conversations, participation demonstrated an awareness of the need to keep their data secure, they talked about the importance of information going to the right place and being shared with the right people and for individuals to take responsibility for keeping their data safe. Some were cautious about utilising online services which they perceived to be highly susceptible to hacking or misuse, while others had confidence in the online security provisions in place, for example by banks, to keep their data safe as long as they were sensible about sharing their data.

> *"I asked a company to help me apply for my tax refund, they charged $20, I had to give them lots of information. I had to do it to get the refund. I had some concerns about giving them my bank details, I thought there was a possibility they could hack into my bank account, or pull the refund to someone else's account, but it worked out well, they were a trustworthy organisation"*
>
> *"My husband was accused of stealing, but someone else was witnessed…and convicted. My husband lost his job over it, there is still a red flag on his police record which makes it hard for him to get a job and it's had a big impact on him. The police should not have shared that information or continued to hold it on his record."*

Inappropriate and / or inaccurate sharing of information on social media was a real concern for some people. While others recognised the benefits of this type of sharing, for example, to offer support or assistance or provide information.

> *"My art work was sold through my art school's Facebook page to raise funds for victims of cyclone Pam. Photographs of my art work with my name were posted by the art class. I felt proud"*

Impact of integrated information going to the right or wrong place

Participants were generally in agreement about the types of organisations that they were comfortable with having access to their data, as well as those organisations that they would definitely not want their data shared with.

Examples of the *right people* to share data with and between included;

- Government organisations (Work and Income, Inland Revenue, Ministries of Justice, Education and Health, Department of Immigration, Housing NZ)
- Government affiliated community services
- Police

- Schools
- Health care providers
- Banks
- Employers
- City council – (specific department only)

Examples of the *wrong people* to share data with included;

- Press
- Social media (by some but not all)
- Insurance companies
- Non-government organisations
- Loan sharks
- Pawn shops
- Internet generally

Impact of data integration on the public at large or organisation

Generally the way people talked about data integration situations suggested they thought it was useful and necessary for data to be shared between government departments and other affiliated organisations as this assisted them to do their job well by providing them with the knowledge to: identify the problems that need to be addressed, prioritise spending, improve service provision, improve efficiency, ensure continuity, inform decisions, and that this in turn would deliver benefits for both individuals and the wider community.

Participants talked about the police being a particular organisation who would benefit from data integration to help them do their job well. For example by having access to integrated information from telephone calls, mobile phone GPS location, CCTV footage, Ministry of Justice proceedings and banking transactions to monitor, or detect criminal activity, or apprehend criminals.

> *“some people need listening to, with phone tapping, the information is all about crime, it is useful and acceptable, it only matters if you have something to hide”*
>
> *“they lose their right to privacy when they become criminals”*

Some participants also felt it would be useful for data to be shared with the police in order for them to be alerted to high risk individuals living in their area, and consequently be able to monitor them in the community, and where appropriate share information with community leaders to promote safety within the community. Although this last point was contentious for many and a large number of respondents felt strongly that information around criminals should remain with the police, and not be shared within the community as this could have negative outcomes for both the community residents, in terms of increasing fear or encouraging a hostile reaction and for the previous criminal in not enabling them to have a fresh start free of prejudice or judgement.

The Department of Internal Affairs was an example discussed where both national and international data integration around health, education, skills, experience, birth and marriage information and criminal records would assist this department to ensure that the most appropriate people are entering the country.

Participants talked about suggested benefits for the Ministry of Education of collating information about students in relation to their level of achievement and progress towards national standards with the aim of identifying schools that need additional support and resources.

The integration of student's health care data with education data was also considered important to ensure that student's individual needs are catered for, such as where a child has special needs that require additional interventions within the school environment such as safety gates or a wheelchair accessible environment.

Participants also talked about integrating information about a child and family to enable schools to identify children and families that would benefit from extra support or a different approach, identify or monitor vulnerable or at risk children, and work together towards solutions.

> *"Sharing the information is good because other people might find a solution for these kids and be willing to help"*
>
> *"I worked for these people who ran a poly-tech, the first time I went there was nobody there, when I asked how many students he needed to run a poly-tech he said about 300, he had taken names from a survey that people had signed down at the mall and used his funding to buy a café. There should have been information to show that these people were not registered to study, The Ministry of Education and whoever provides the funding did not do their checks in time. They must have known there was no data sharing as they were not worried at all about it"*

2.5 Acceptability

Theme: Trade-offs that make data integration acceptable

The participants talked about a range of factors that would make sharing their data more acceptable, the most salient being restrictions on who used the data and for what purpose. The bottom line that the majority of participants talked about was that data integration should only be used to have a positive and/or fair outcome either for the individual or community at large. Through their conversations participants talked about a range of values underlying their attitude towards data integration. As a whole they valued positive purposes and outcomes of data integration, fairness in its application and the individual taking responsibility for both keeping their data safe and accepting the outcomes of data integration.

Acceptability factors

A number of method or system factors that would make data integration acceptable came up repeatedly in participant narratives, including the need to

- Restrict who information is shared with to only those who need it
- Restrict who information is used by to only trusted organisations
- Restrict what information is shared to only that which is relevant
- Do not share widely
- Do not share with private organisations
- Ensure anonymity for statistical research
- Use only for individual or wider public benefit, or to provide value
- Seek permission to use data

- Ensure fairness when using data to access public funds, resources and benefits
- Be sensitive about how information is shared, who it is shared with, when it is shared
- Maintain privacy, prevent prejudice
- Ensure the needs of the individual are balanced against the overall benefit to wider society

Values
The way in which participants talked about positive aspects of data integration indicate a number of key values that were important to them, including

- Benefit to community should overweigh benefit for one individual
- Public good is more important than individual right to privacy
- Individuals should take responsibility for consequences of their actions
- Importance of confidentiality and protection of privacy
- Permission should be sought
- Organisations using data should have public interests as their focus
- Not to be used by private organisations to profit, to target, exploit or judge
- Data should not be misused
- Needs to be real value from integrating the data
- Data doesn't capture the whole picture about a person

Bottom line
Overall, there was recognition that trade-offs were necessary but only acceptable if certain factors and principles were maintained, such as:

- Use only for benefit, to improve things, to make life easier, to provide solutions
- Share only with the right people for the right purpose
- Do not expose people or make them vulnerable to targeting or exploitation
- Outcomes of data integration should be fair for all
- Responsible use
- Appropriate use
- Restricted access

Theme: Barriers to the acceptable use of data integration

While many of the participants recognised the value of technology in assisting data integration processes, there were also some for whom technology was seen as a barrier because of perceived security issues and potential for misuse, particularly in relation to online processes. Another barrier talked about was the possibility of information not being kept safe and ending up in the wrong hands, again with the possibility of the information being misused as a negative outcome.

Safety as a barrier to data integration
While the participants generally trusted government organisations who hold data to keep their data safe, some also had concerns around the safety of their data, for example;

- The possibility of information ending up in the wrong hands
- The possibility of data sharing between government organisations and private organisations
- The possibility of organisations holding data being bought up, merged or taken over by companies with different values and priorities

> *"You wouldn't want a loan shark setting up in an area where there is a high level of poverty, people won't turn down a pot of gold, but it will be more detrimental to them in the long run"*
>
> *"Its not so useful or acceptable (sharing health and lifestyle information with insurance companies), as far as I know it doesn't matter what the person looks like, whether they are fat or thin, you have to treat everyone the same"*

Technology as a barrier to data integration
While respondents recognised the value of technology in delivering behind the scenes data integration processes, they also had some concerns around possible misuse of data through technology. For example;

- The use of GPS (global positioning systems) for surveillance
- Questionable internet security for online banking, shopping, and online activity
- The possible misuse of Facebook providing access to personal, friends and family data
- Implications of Pay-Wave contact free banking transactions

> *"I don't trust the internet, I don't know if it's completely safe, people can pick stuff off it, they could use my information to create a fake passport"*

Competing risks to individual versus public versus organisation as a barrier to data integration
The need to balance the risk to individuals over wider interests was discussed in depth across the narrative interviews. The over-riding feeling emerging around risk to the individual was that while it was acceptable and necessary for official processes to be completed, that the method and timing of these interventions should be sensitive to the circumstances of the individual concerned, to minimise any negative impact of the data integration.

> *"it is fair (paying tax on 2nd part-time income), but I ended up getting a bill and owing them money, I wish they had paid me less up-front instead of sending me a bill later, if the tax had already been taken care of before I was paid, it would have been less of a kick in the teeth"*

For individuals with a criminal record, for example, the respondents tended to prioritise the safety of the wider community over the possible costs to the individual of difficulties relocating and seeking employment. They were however, generally appreciative of the difficulties around the boot being on the other foot, and addressed this by proposing that information about a person's criminal past should be restricted for sharing with the police and law enforcement agencies only, so that while the police were empowered to protect the community, the individual concerned was afforded a degree of privacy to have the possibility of a fresh start.

> *"They could be trying to get a new start, moving away from their old neighbourhood where all their crime buddies are, but then get smacked down before they start"*
>
> *"Everyone is entitled to a second chance in life, but my priority is with my community and personal safety for my family"*

This type of response was also evident around the provided scenario where information is shared about a recently widowed woman. Participants recognised the need for official processes to take place that might have a negative impact on the widow, for example, a financial penalty through change in her entitlements, but reconciled this by suggesting that the timing of the communications around this could be more sensitive to her immediate grief.

> *"they could make a phone call or something, ease her into it if they know it's going to be a big knock"*

They also recognised that while targeted marketing could be intrusive and insensitive in the early days, it might also be useful in providing information if delivered at an appropriate time.

> *"at some point it could be useful to her…she might be interested in a couple of years' time, but it's not appropriate now"*

There was, however a strong feeling among some respondents that if an individual had deliberately attempted to deceive a government department for their own benefit then they should expect there to be a consequence when data integration processes revealed their discrepancy. The element of fairness coming through strongly as a necessary quality of data integration outcomes.

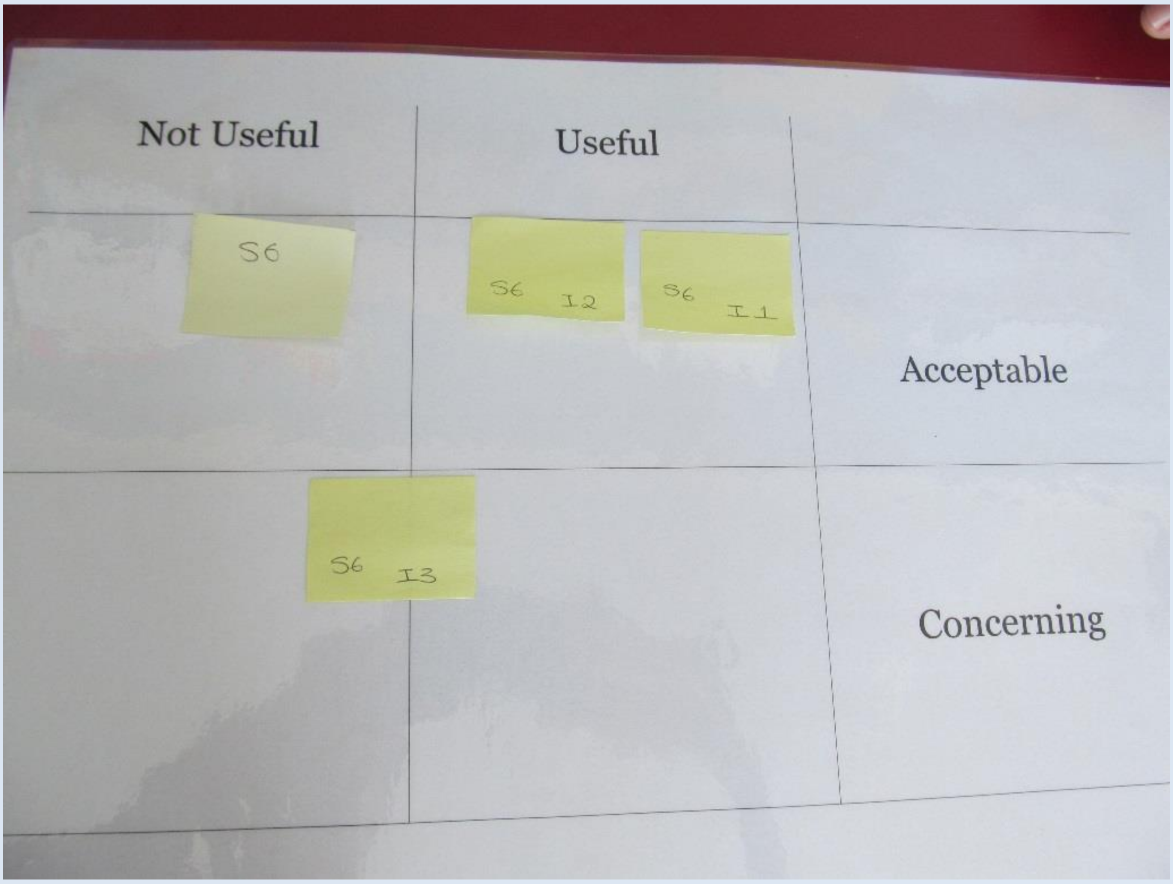
> *"I don't mind them talking to each other (IRD and WINZ), she knew what she was doing, she did it to herself and she got caught out, it is what it is"*
>
> *"it should happen more (detecting benefit fraud), people use different names, an English name and an Island name and can claim for different children using different names, they (WINZ), might have to ask for more detailed information about them at the beginning"*

Lack of trust as a barrier to data integration
Many of the participants talked about how there needed to be trust in organisations to use their data appropriately, and for the intended purpose.

2.6 A highlight example of how acceptability can shift

The photo in Figure 1 shows an example of a participants responses for scenario 6, which introduces the linking of information to profile specific groups of people. Initially the participant felt that this type of data integration was acceptable but not useful. However, with the introduction of the first two interventions around the council and local businesses using the information to plan and target services respectively, the participant then amended their view to indicate they felt that this would be a useful purpose for the data integration. Then when the third intervention around loan sharks having access to the data was presented, the participant felt that this was concerning, as illustrated by their positioning of the post-it note on the chart.

Figure 1. Usefulness/Acceptability Photo from Narrative Interviews



2.7 Statistics New Zealand as custodian

Theme: Attitude around the custodians of data

The participants expressed a general level of trust in Statistics New Zealand to act as custodians of their data, to place restrictions on who can access the data and on how much information data they can have access to. The participants felt that confidentiality would be maintained and there was an expectation that Statistics New Zealand would only provide information to trusted organisations, typically government departments, who would use it for good reason, or a useful purpose.

2.8 Operational vs Statistical Use

The participants talked about a wide range of experiences around operational use of their data, and on the whole they considered that their data had been used appropriately and in a useful way. A few participants also talked about dissatisfaction with the process or outcomes of data integration.

There was little in the interviews to suggest that participants had particular or different concerns about the use of their data for statistical purposes where the data was anonymised and pooled with other data. That is, participants identified the same need for confidentiality and purpose for all data integration use.

There was an expectation and general level of trust that Government organisations would use statistical data integration for good use, for example, to inform health promotion campaigns, to improve service provision, expand knowledge base, or identify need.

> *"I would be happy for the information to be used if it is going to help other parents… if it would benefit my child. I trust the government to look after the information.*

A small number of participants were concerned about possible misuse or negative outcomes of statistical data by organisations that they considered be in a position of power, for example, by using the information to discriminate against particular individuals or communities.

The main area of concern raised in interviews was about the use of ethnicity data in ways participants considered unhelpful. For example stereotyping, making judgements about or profiling communities was of particular concern in terms of "painting a negative picture that would affect how community residents feel about themselves".

2.9 Summary of common themes and differences across groups.

A range of operational and statistical examples were covered in participants' narratives. The operational examples provided by the participants were varied, but interactions with health care providers, the education system, Inland Revenue (IRD), Work and Income, (WINZ) and the justice system were most frequently described.

The key themes emerging from the narrative interviews were that:

- Data integration is over-all considered by the participants to be a useful, beneficial and positive thing, but does also have the potential for harmful outcomes if data is misused;

- Participants considered it extremely important that data is shared with and used by the right people and for the right purpose – to provide benefit, rather than harm. They have a high degree of trust in Statistics New Zealand as custodians of data, and in Government departments to use their data appropriately and primarily for good purpose;

- Issues about consent to share information were often discussed. Concerns were expressed about private organisations or third parties having access to data which could then be used for the purpose of profiting or commercial gain, to target or exploit vulnerable people. They felt strongly that information should not be shared with non-Government or private organisations without their consent/permission;

- Participants felt that consideration should be given to the specific context of the situation when deciding who needs to have access to their information, how much they need to know and what is relevant, for example whether historic information should be shared or not. It was generally considered that in most situations, wide blanket sharing of all information is neither necessary nor acceptable;

- Participants voiced expectations around data integration being used fairly, that outcomes would be applied equally to everyone and that individuals should expect to take responsibility for any negative consequences of data integration due to their previous actions, e.g. a criminal record, or benefit fraud;

- Statistical data integration was considered by all as definitely a good thing that provides organisations with access to information which enables them to do their job well.

Anonymisation of data through de-personalising of pooled statistical data was not seen as problematic, as long as it was used by the right people and was used to have a positive impact for either a particular group of individuals or wider society. There was a high degree of trust in Government organisations to use integrated data to inform decision-making, provide benefits or improve services.

The interviewees had no concerns with Statistics NZ being the *guardians* of the data, and were happy for their individual information to be used as part of a group, for the purpose of informing Government departments or undertaking research. Their criteria for data use of this type was that there was no unique information identifying them individually, and that the data was being used for the positive benefit of the wider community / population concerned.

2.10 Key differences between groups

We did not find noticeable differences in attitudes to data integration between the target population groups. As would be expected, there were attitudinal differences observed between individual participants. Across all groups there were many similarities, particularly in relation to the core values and themes. The majority of variation in attitudes and perceptions stemmed from the participants own personal experiences and where they could personally relate to all or some aspects of the scenarios, rather than being attributable to a particular group as a whole.

3 Results: Workshops

Four workshops were conducted, one in Invercargill, one in Gisborne, and two in Wellington. The workshop method was developed with the intention of un-teasing complexity, by direct observation of participant reaction to data integration using concrete manipulatives (children's building blocks) to represent data fields. Concrete manipulatives enable participants to link information provided during the workshop with personal opinions and knowledge in a meaningful way. The plastic blocks were identical in size and shape and coloured coded to represent the government agency for which the data was originally collected, with the data field written on each block. The data fields were personalised, for example "my relationship status" rather than "relationship status" to emphasis the linking of data about each participant. See Appendix I for a more detailed methodology.

3.1 Participant characteristics

Thirty people participated in total. The participants' demographic characteristics are shown in table 4 below.

Table 4: Workshop Participant Characteristics

Gender	Per cent	Age group	Per cent	Ethnicity	Per cent	Occupation	Per cent
Female	47	Under 25	7	NZ European	67	Paid employment	37
Male	53	25 to 34	20	Māori	30	Self employed	3
		35 to 44	13	Pacific Islander	27	Unemployed	17
		45 to 64	16	Other	0	Retired	20
		65+	23			Other	14

Note: Not all categories add to 100 per cent. Some data was missing and sometimes more than one category was selected.

3.2 Responsiveness to the workshop instrument

The research instrument (moving data field blocks) engaged the participants in the reality of data integration very well. During the workshops participants could choose whether or not to physically move their personal information 'block' from a range of government agencies into Statistics NZ, which allowed them to fully appreciate the characteristics, process and implications of data integration. There was a high level of expressive involvement and discussion amongst participants in the workshops, and many participants commented that they had found the experience enjoyable. Some commented that they hoped there would be further opportunities for public consultation on the topic.

3.3 Findings

The action research workshops found that statistical data integration acceptability was dependent on a wide range of considerations. These include:

- The type of data
- The need for and value of data integration
- Personal experiences including stereotyping and profiling
- Capturing meaningful information
- Confidence in Statistics NZ as a custodian
- Privacy
- Security and surveillance
- Transparency and consent

Each of these considerations are now discussed in more detail, with participant quotes employed to highlight particular points of view.

3.3.1 Data type

Information sharing acceptability did not seem to be greatly affected by the *source* of the information, but it was considerably affected by the *type* of information.

Participants in the workshops were generally happy with sharing data that is already commonly used for statistical purposes, such as:

- My birth date
- If I've ever been on a benefit
- My address history
- Whether I'm a smoker (or drinker)
- My migration status
- How often and when I enter and/or leave NZ
- Medical Conditions I've had
- My ethnicity

These data were seen as being routine, commonly collected types of information, and were not seen as being particularly sensitive. Typical comments for these data types included:

> *"Yeah that's fine, they know that anyway"*
> *"It's nothing too personal so I don't really care"*
> *"Yeah, that's good data to have"*
> *"They should just all be in the stats"*
> *"Everyone already has that"*
> *"[Address history data] would be quite interesting actually"*
> *"There's no feedback [to other departments], so it doesn't matter"*

On the other hand, either there was no full consensus, or some or many the workshop participants were not comfortable with sharing:

- My history of interaction with the Police
- My family's death records
- My family's birth records
- My marriage records
- How many and the type of allowances I've claimed
- My Partnership history and/or status

- My history of convictions/ My sentence history
- My educational history
- The amount and repayment of my student loan
- Number of visits to Hospital or GP
- My exact income every year
- Whether I pay or owe any child support

In general, workshop participants were less comfortable with sharing these types of information as it was seen as being too personal, complex, or sensitive, or they were not convinced that it *needed* to be shared: "*why do they need that? No, that's too much information*". For example, many people felt that sharing 'my history of interaction with the police' with Statistics NZ at an individual level was going too far and unnecessary:

> *"I don't think people who had that would like that to be shared knowledge."*
>
> *"I'm sure they need to know statistics on interaction with the police, but personally I wouldn't want my information to be in Statistics NZ."*

Both the ONS and the Office of the Australian Information Commission (cited by Sherriff 2013) similarly found that some types of data, such as income and health information were considered more sensitive, and that people were uncomfortable with having this type of data shared.

Some participants felt that Police information on peoples' convictions or sentence history was okay to share, while others felt strongly that it was too personal and inappropriate to share beyond the police databases.

> *"If you have a conviction it's pretty much public information anyway."*
>
> *"I want criminals' information to be in there but not mine. That's fair enough, you do the crime you do the time."*
>
> *"How far in the past does the sharing need to go? Some may be so old that it's no longer relevant. That shouldn't be shared if it's old, minor convictions."*
>
> *"How's my criminal history relevant?"*
>
> *"Yeah so it's police information, why does it need to be statistical information? If police already have it, and they make their decisions based on that, why does it need to go to Stats, it's not to do with them"*

Others were interested in how information such as interaction with the police and conviction/sentence history may be used, highlighting that it could be used for better or worse:

> *"But then it might bring help from the government."*
>
> *"Or it could just make the police clampdown."*

Others felt that sharing some types of information was going too far, as it was very personal and they could not see how it was useful information to use statistically, such as peoples' relationship histories.

> *"What's [my relationship history] going to help with? I don't see how that's going to help."*

However, about a third of the workshop participants had no reservations about sharing all their data with Statistics NZ.

> *"What harms in that, I can't see it. Put the lot in."*

While the type of data being shared with Statistics NZ was particularly important, there were a range of other considerations which also affected peoples' acceptability of statistical data integration.

3.3.2 The need for and value of data integration

People were highly inquisitive about why there is a need for data integration, the value of data integration and what the integrated data would be used for, rather than just the 'dry' exercise of integration in and of itself.

The participants asked questions such as:

> *"I want to know who's holding my information, why they've got it, what they're going to do with it, and how secure is it'*
>
> *"Can you please explain what StatsNZ are going to do with this?"*
>
> *"So this is all linked on my name?"*

Some participants felt that data integration was not necessary, as statistics are already reported on by government agencies.

> *"Well where do they get this info from anyway? Don't MSD release their own stats on this? So if they're already releasing the statistics what's the point of Stats NZ having it?*
>
> *"If they need stats they can come [to this ministry] and get it, I don't think it should be shared as a matter of course."*
>
> *"If Stats NZ want to know information, they can ask the ministries."*

Others felt that integrating data may help to create a powerful database which could reveal trends and relationships that are currently unknown or missed. It could also help with forward planning such as levels of funding for health care in the future. For some participants, if they could see how the information could be used to improve fairness and peoples' wellbeing, and if they thought it was in the public interest, then they tended to be more supportive of data integration.

> *"A lot of people just get thrown back into the workforce, and if stats are looking for that it could reveal things that aren't working."*

> *"Statistical integration may help identify gaps that otherwise wouldn't be found, which can inform ministry policies."*

> *"If they're the organisation which gives information and feedback to the government about what's going on in the areas, it should be as accurate as possible… cos if the government doesn't have all the data, or the right data, they're not going to hit the nail on the head."*

> *"There's a lot of information in there that will help out a lot of people"*

Some workshop participants felt that integrated data could be more accurate, reliable, and up to date than existing data, particularly if existing data was self-reported.

> *"You put rubbish in, you get rubbish out"*

> *"We know current statistics aren't very good anyway, so this way they may be better."*

Others felt that some of the government agencies should be sharing information with each other, but they should bypass Statistics NZ altogether. Others pointed out in response that looking at complex statistics may help to point to gaps where data needs to be shared between ministries.

Some participants thought that data integration was an efficient and natural progression with the possibilities created by technology today.

> *"50 years down the track, looking back things will have changed so much, and this is just a progression which makes sense."*

> *"It makes sense cos each of these departments have got segments or portions of our history and background, why can't it all be in one place? If you have problems with big brother you may not like it. But some of the data that's held by these departments, I wouldn't be sure that it hasn't already been shared, so to my mind this could be going on already but perhaps there isn't, and I'm just assuming it is."*

> *"Well then you wouldn't need to fill out a census, and it would be more accurate than self-reported census data every five years, save time and money"*

> *"In years to come, a paper survey will be seen as a pretty crude way to take stock of the country."*

Some participants questioned whether the benefits of statistical data integration could justify the costs:

> *"How much is it going to cost to set up these systems, versus the benefit? I see that as a massive task".*

3.3.3 Personal experiences with statistics: stereotyping and profiling

Consensus of opinion was not completely reached within each workshop group, nor is it overly evident within each of the population groups of interest. Rather, it appears that people's attitudes to data integration are best attributed to their personal experiences with government and statistics, as well as the types of statistical categories people's lives fit into and how these may have been stigmatised in the past. This included the categories of age, education, occupation, ethnicity, use of allowances and benefits, criminal history and area of residence. Some people expressed frustration at feeling constantly negatively stereotyped or profiled based on these statistical categories, and felt that statistical data integration could just make things worse.

> *"What gets me is one of these blocks is nothing, you don't worry about one particular thing, but you group that together, you're grouping that into one particular person and going, well, that's a bad person, that's a good person. That's the danger of having all this information in one place."*
>
> *"I know [my town] has a bad rate [of crime], and that would be painting [my town] in a bad way, when a lot of people there are really good."*
>
> *"That's what it comes down to, stereotyping certain families."*
>
> *"Whether it's okay or not depends on one's experiences"*
>
> *"We don't want to be getting into racial profiling or anything like that, that's not cool. I think it will increase profiling, especially if that information is going to be shared."*

Others felt that it could help to dispel negative stereotypes, by showing them as not true or demonstrating change over time. They noted that this would depend on what data is collected and how it is analysed:

> *"Everything's already stereotyped already anyway right, so why not get it for real?"*

3.3.4 Capturing meaningful information

Related to people's experience of data being used to reinforce stereotypes was discussion around whether the data types used in the workshop were meaningful. Some participants raised that they felt that common statistical data does not provide a good indication of how well we are truly doing as people, and that the wider context for statistical data is often missed. Some also felt that statistical data, and the categories used within the data, may not be a true, fair and inclusive representation of a diverse reality. For example, some categories that people feel they belong to may not be an 'option' in the data, such as their ethnicity or their sexual orientation. Lastly, some people felt that some types of data were not objective and reflected biases or prejudice in the organisations collecting the data, such as racial prejudice in the Police. Integrating this biased data could then lead to incorrect conclusions being made about certain populations of people.

For these reasons, some people felt that a risk of data integration is that it could simply provide a larger amount of not very useful and not very meaningful information:

> *"Most of this stuff is pretty deficit based… there's not a whole lot that gives you parts of the person that's great, what's here that's positive? I think the thing that's missing here is all this data doesn't tell you the whole story, it gives you a bunch of little indicators, but it doesn't tell you about the people."*

> *"Yeah like it doesn't tell you why people are on the benefit, all this data misses the why, the context. Just looking at the measures they're using, I don't know how you can get to a why when the indicators are so black and white. "I mean you could say you're Māori, you're on the benefit, under 25 with a baby, but there's a bigger picture than that… but they're just getting stereotyped."*

> *"They're asking the wrong questions, and using the wrong data in the first place"*

> *"Why don't they ask things like how many neighbours do you know, what interaction do you have within your community? Look at how we're living rather than just the black and white."*

> *"If the data they used was more meaningful and focused on peoples' wellbeing then people may be more engaged in statistics and want to fill out the forms. If they ask nice questions, I want to give a nice real answer, and hopefully they're going to use that to make things better."*

> *"Brown people get stopped by Police more often, and so they get incarcerated more often, the way information is being used now is pretty damaging."*

> *"I don't know if income every year is a true reflection of where you're at."*

These comments highlight the perception among some participants that perhaps the types of data commonly used to describe the population are inadequate, and investing in a larger volume of more detailed, but inherently inadequate data could come at the expense of capturing more diverse and perhaps more meaningful data about how well people are truly doing.

3.3.5 Confidence in Statistics NZ as a custodian

Some of the workshop participants were trusting of Statistics NZ as a custodian of integrated data and some were not.

> *"I don't have a problem bearing in mind it's going to Statistics and it's not shared, and no one else gets access to it."*

> *"So I'd be worried if this information was sold, to a private company, if it's in Stats NZ I'm comfortable, but if it was sold to Mr Google then I'd be very hesitant."*

> *"Probably Statistics is the only government department that can be depended upon to be completely independent and relevant. They're just gathering information, that's all they do. It's not IRD or ACC. They're less controversial."*

> *"I have no doubts about the integrity of Stats NZ information"*

> *"Statistics NZ need to change their name, and it may change their attitude, and portray a better, a more engaging type of response. Statistics to me are scary, but if they could encompass wellbeing, I'm sure people would be more responsive."*

> *"Whether it's okay [for Statistics NZ to have the data] depends on the function that Stats NZ provides."*

3.3.6 Privacy

When workshop participants discussed integrated data being held in a single database and linked to their either their name or other personal/identifying information they tended to become increasingly less comfortable with statistical data integration. Some people felt that the information was best retained and analysed by the agency that held it, with that agency providing collective statistics only.

> *"I don't mind the information going, I just don't want my name to go with it" (general round of agreement to that)" / "Yes but that applies to all that, to the whole lot."*
>
> *"I see no reason to bog this down with the individual aspect, I just see absolutely no reason."*
>
> *"That's fine, but my name don't need to go with it, anything can be in there that they want, but not my names." / "Yes, no names at all for anything." / "Yes you're looking for group responses, not individual." / "Yes but then your information could be in six places instead of just being in one, and they can't link it cos they don't know who it belongs to."/ "But it doesn't matter cos they're getting collective from here, here, here, here, and it puts an onus back on these other departments to make sure it's tight, secure, and correctly accumulated.*
>
> *"My only question is, why does Stats NZ need personal information? They don't need to know my personal information."*
>
> *"The important thing is how is this confidential information treated? Is it strictly confidential and anonymous or can it be misused?"*

3.3.7 Security and surveillance

Many of the workshop participants were concerned with security risks and, to a lesser extent, the misuse of integrated databases for surveillance.

> *"If everything here is in that Stats NZ about me, if someone hacked it, they would know everything about me, and that's scary, that would have to be so secure, otherwise it's too much information they've got about individuals, it's very powerful."*
>
> *"How competent are the people handling all this data, and I always think of Novapay, and there it was a lack of skill of people who were managing the system, and will the system be audited for quality?"*
>
> *"New Zealand has a good reputation, I believe for safeguarding privacy and information, unless you're ACC"*
>
> *"The IRD are a good example of success in this area, cos they've exported their system, but they're the only one"*
>
> *"They're not going to be secure, they're going to get hacked."*

> *"I'm always going to be suspicious about big brother, no matter what we de he's watching."*

> *"If the data is kept individually in each department, it's less likely to be interfered with. Collective info can go to the stats department, but I don't see a need either for individual data linked to a person."*

3.3.8 Transparency and consent

Some participants felt that the public should be fully informed about data integration practices, as well as have the choice as to whether their information was integrated or not:

> *"I think informed consent is a really big one, people need to know what's being shared and how it's being used, and why they need that, and that you can say no to having your data integrated."*

3.4 Workshop summary

The action research workshops found that statistical data integration acceptability was dependant on a wide range of considerations.

Information sharing acceptability did not seem to be greatly affected by the source of the information, but it was considerably affected by the type of information. Participants were generally happy with sharing less personal data that is already commonly used for statistical purposes. In general, workshop participants were less comfortable with sharing information they felt to be personal, complex, or sensitive, or if they were not convinced that it was necessary to share it. People were highly inquisitive about why there was a need for data integration and what the integrated data would be used for. If participants trusted that data integration would be used to produce public benefits and increase fairness then they were much more likely to support it. If they felt that it was unnecessary, not useful, could be misused, or was too costly they were less likely to be supportive of it.

It appears that people's attitudes to data integration can be strongly attributed to their personal experiences with, and trust in, government authorities and statistics, as well as the types of statistical categories people's lives fit into and how these may have been stigmatised in the past. Related to people's experience of data being used to reinforce stereotypes was discussion around whether the data types used in the workshop were meaningful. Some participants raised that they felt that common statistical data does not provide a good indication of how well we are truly doing as people, and that the wider context for statistical data is often missed. Some also felt that statistical data, and the categories used within the data, may not be a true, fair and inclusive representation of a diverse reality. Lastly, some people felt that some types of data were not objective and reflected biases or prejudice in the organisations collecting the data. Integrating this biased data could then lead to incorrect and unfair conclusions being made about certain populations of people.

Many of the participants were concerned about the potential for security risks and, to a lesser extent, the misuse of integrated databases for surveillance. Some of the workshop participants were trusting of Statistics NZ as a custodian of integrated data and some were not. When workshop participants discussed integrated data being held in a single database and linked to their personal or identifying information they tended to become increasingly less comfortable with statistical data integration. Some people felt that the information was best retained and analysed by the agency that held it, with that agency providing collective statistics only. There were some participants who felt that data integration was only acceptable if peoples' informed consent had been given.

3.4.1 Data Integration and the Census

Of particular relevance to the Census, was the perception that bringing in administrative data from other sources was seen positively, in terms of accuracy, efficient government spending, and valuing public time.

> *"Well then you wouldn't need to fill out a census, and it would be more accurate than self-reported census data every five years, save time and money"*

Also, the theme around public wellbeing was raised as an important one around public engagement and improving the quality of the data and improving public response to the Census generally.

> *"If the data they used was more meaningful and focused on peoples' wellbeing then people may be more engaged in statistics and want to fill out the forms. If they ask nice questions, I want to give a nice real answer, and hopefully they're going to use that to make things better."*

4 Results: Online Consultation

The online consultation adopted a modified Delphi approach to explore concerns with a selection of three controversial data integration scenarios amongst a range of experts. Then, to address the concerns, experts investigated potential constraints on how the data was used, and benefits associated with the data integration scenarios to see if this would alter acceptability. The Delphi Technique is a method of structured communication in a workshop setting in which a panel of experts answer questions in two or more rounds. First the questions are answered individually, then discussion and further rounds lead to a group consensus (Hsu & Sandford, 2007). See Appendix I for a more detailed methodology.

4.1 Participant characteristics

In total, 43 subject matter experts were invited to participate in the research. The final group of 15 people that indicated they wished to participate in the online consultation included academics, journalists, technology experts, city council staff, and members of pertinent government agencies. However, only 13 people participated in the online consultation.

Participation in each phase of the online consultation varied, with the greatest number of people contributing to the first phase, which explored concerns with the three scenarios. Participation was lowest for the second phase, in which the experts were asked to contribute ideas regarding constrains on the scenarios that would help to decrease or mitigate concerns highlighted in the first phase. The final phase investigated benefits associated with the scenarios and had slightly greater participation than the second phase for both scenario 1 and scenario 2. The number of contributors to each phase and scenario is shown in Table 5.

Table 5: Number of participants that contributed to each phase of the online consultation

	Scenario 1	Scenario 2	Scenario 3
Phase 1: Concerns	13	11	10
Phase 2: Constraints	4	5	5
Phase 3: Benefits	7	7	5

4.2 Participant response to the online consultation

The full potential of the online consultation with experts was not realised. Whilst interest was initially sparked in the topic, the consultation did not maintain adequate momentum or energy to sustain the same level of participation in subsequent phases. This behaviour might be explained by a number of converging factors:

- the recruitment effort was constrained to a short timeframe, with little lead-in prior to the online consultation
- the long weekend between phases 1 and 2 may have placed pressure on participants to complete work during the short week that phases 2 and 3 were conducted, leaving less time and energy to contribute to the online consultation
- the initial phase may have been more engaging because the consultation and topics were new
- it may have proved easier to contribute concerns rather than constraints and benefits
- the constraints and benefits phases may have been less engaging because the scenarios were controversial and unrealistic
- if participants had readily cogenerated (edited) the ideas already contributed, there may have been a greater sense of participation and group ownership leading to greater participation in the later phases
- similarly, cogeneration may have improved the quality of contributions such that each contribution would only represent one idea, thereby making the ranking exercise more fruitful, encouraging greater participation
- the anonymity of the consultation may have worked to discourage participation
- the trade-off in having cogeneration and ranking features in Codigital made the contribution format rigid and difficult to moderate; a more interactive format may have encouraged greater participation.

The following comments were captured during phone calls with participants near the end of the online consultation to encourage participation in the last two phases (most attempts to ring participants were unsuccessful):

- One academic commented that they “found it straight forward to take part, it didn’t take long”. The tight timeframe for participating in each stage kept things running. However, “it was not as interactive as I’d expected”. This person looked at all three scenarios but chose to focus on one in particular
- Another participant thought the approach was good and enjoyable but that the scenarios were more like TV than real life. Also felt the tight timescale for comments was good to stop some people “waffling on”, but was too busy to provide much input.

Also felt people had missed the point a bit but didn't take the opportunity to reword statements. Felt there was a need for more active moderation, and that if there was a centralised database "it would be used for issuing parking tickets, not catching criminals"

- A government employee commented that the process was straightforward but felt it was impossible to overcome or address the concerns. This respondent thought that the scenarios were too extreme and unworkable, so it was a waste of time to even think about it.

Of the six people that participated in the post-consultation survey, five indicated that they felt they had enough of an opportunity to contribute to the consultation and the other person indicated that they were very busy at the time and were unable to contribute as much as they would have liked. They contributed the following comments about the online consultation process:

- *"interesting process. good luck :)"*
- *"The suggestions were too extreme and too simplistic to be modified into workable options."*
- *"2 days was a tight time-frame when a lot of people didn't put in comments until late on the second day thus there were only comments and votes on the early points. Still have no idea how "your contribution" rank and percentages were calculated. Seemed slightly strange. Felt very hesitant to edit other people's comments very much. Easier to put in a new comment. If this had been done in a face-to-face meeting, we would have built on the comments better and probably got one overall paragraph."*
- *"There were two types of concerns that people raised: privacy concerns and concerns about whether the idea was a good one. For example, I have very little problem, in terms of privacy, in using administrative data to model what students do well in what courses. But I did think it was a very bad idea to base course fees or admissions on such a model. It would have been useful to separate these two concerns and discuss them separately. For the record, my answers earlier in this survey mostly reflect concerns about whether the ideas are good ones, not privacy concerns."*

4.3 Findings

4.3.1 Scenario 1: A single view of citizen costs

Scenario: *What if we could use a large integrated dataset to measure how much each person is costing their fellow citizens/taxpayers? We could use information from agencies such as Ministry of Social Development, Justice and IRD to track and/or predict the cost of services over time. We could also see if people use more government support than their taxes cover.*

Participation in the online consultation was greatest for this scenario, with all 13 participants making contributions.

4.3.1.1 Concerns

Participants were asked: *What are your concerns with the following data integration scenario?* The concerns raised in scenario 1 are shown in Appendix IV, Table 11. These can be summarised as:

- that individuals or groups would be stigmatised
- that it restricts the value of individuals to financial performance alone
- that it would be unfair
- that it does not represent complex relationships
- that it indicates a one to one relationship between tax contribution and cost to society, which is inaccurate
- that it would create an underclass
- that it would lead to strategic damage because the assumptions are flawed

Further concerns were contributed in subsequent phases relating to this scenario. These concerns can be summarised as:

- the information will not help government departments target interventions
- it would be incorrectly viewed as a person's cost to society, leading to social tension
- it overrides the purpose of a social safety net, which is to redistribute resources to those in need
- there is a risk we lose the wider public value of investing in public services
- once individual data exists, users will demand this data even when the data are not permitted to be used in a non-aggregated way

4.3.1.2 Constraints

During the online consultation participants were asked: *What sort of safeguards, constraints, or changes would you add to the following scenario to mitigate your concerns*. Only four people contributed ideas, which are shown in Appendix V, Table 12. These can be summarised as:

- develop broad indicators of fiscal cost, rather than calculating cost at the individual level
- remove comparison of government support with taxes paid
- do not share the findings publicly
- include other non-market/intangible measures of public benefit

4.3.1.3 Benefits

Participants were asked: *Assuming appropriate controls were in place to mitigate your concerns, what do you consider to be the benefits of the following scenario?* Seven participants contributed to this phase for scenario 1. Of the six ideas contributed, only three addressed actual benefits (though one contribution included three benefits). The benefits contributed are presented in Appendix V, Table 13. These can be summarised as being useful for:

- broadening the view of policy analysts, leading to assessments with fuller information
- identifying vulnerable population groups
- finding early indicators of future use of government services
- looking for areas where early interventions could lead to future fiscal savings
- defending programs that target vulnerable populations by illustrating their potential benefits.

4.3.2 Scenario 2: An alternative census

Scenario: *What if we replaced Census data with data from administrative sources? You wouldn't need to fill in a Census form anymore. We would know all about you, all of the time, by using information from your interactions with government departments like Inland Revenue, MSD, or the Ministry of Health. We would use this information to create Official Statistics on a monthly, quarterly or annual basis instead of every five years.*

4.3.2.1 Concerns

Participants were asked: *What are your concerns with the following data integration scenario?* Eleven participants contributed to this consultation phase, but there was no cocreation of ideas expressed in this phase of the study. The concerns with an alternative census are presented in Appendix V, Table 14 and can be summarised as:

- inaccurate in comparison to a periodic census because some important data will not be captured/collected and some groups will be underrepresented because they have little/no interaction with government departments
- risky to rely on probabilistic data
- little control over data collection practices
- low public acceptability
- data collected for one purpose needs careful construction to be used for another
- lack of data protections
- it is not clear who can use it and for what purpose

4.3.2.2 Constraints

Participants were asked: *What sort of safeguards, constraints, or changes would you add to the following scenario to mitigate your concerns*? Five participants made contributions to this phase of consultation Appendix V, Table 15, which can be summarised as:

- test the IDI[1] against the periodic Population Census to identify parts of the Population Census that could be shortened or removed to reduce cost and respondent burden
- keep working on IDI as the frame for an administrative census but test it rigorously against the five-yearly census, using at least 2 time points. This testing phase is vital to ensure that it is both fit for purpose and that
- ensure there is broad public awareness and acceptance of how administrative data are being used
- recognise that it takes a long time and large budget to get the right systems in place to adopt an administrative census

A suggestion made elsewhere in the consultation regarding scenario 2 is:

- a central register of dwellings would be necessary for an administrative census to replace the periodic Population Census

---

[1] Statistics New Zealand's Integrated Data Infrastructure (IDI) is the database in which datasets are integrated. Access to the IDI is granted to approved researchers at an aggregated (anonymous) level.

4.3.2.3 Benefits

Participants were asked: *Assuming appropriate controls were in place to mitigate your concerns, what do you consider to be the benefits of the following scenario?* Seven participants contributed to this phase of the consultation. These are presented in Appendix V, Table 16. The benefits identified can be summarised as:

- provision of longitudinal data
- detailed population data at a geographic level at a greater frequency than five-yearly
- potential for service population information
- endless possibilities for research, policy, and interventions
- potential to be more comprehensive than the periodic Population Census

Benefits expressed in other phases of the consultation in relation to Scenario 2 include:

- it could overcome some of the problems with the periodic Population Census, such as capturing people that do not fill out the forms and data inaccurately recorded on the forms
- it would be useful to have some data on a more regular basis than five-yearly to give input onto strategic planning

4.3.3 Scenario 3: Targeted education

Scenario: *What if we used an integrated dataset to predict the likelihood that you'd be able to get a job in your chosen field? The statistics produced would mean we could then charge variable costs for education to manage demand for courses, or to limit course numbers.*

4.3.3.1 Concerns

Participants were asked: *What are your concerns with the following data integration scenario?* Ten participants contributed to this phase of the consultation. The concerns contributed are shown in Appendix V, Table 17 and can be summarised as:

- it may not align with the market demand for skills
- it assumes people are only influenced by earning potential in their study choices
- it would reinforce barriers to education and employment for particular groups
- it is difficult to predict the future
- it is difficult to predict multiple career paths
- the model would skew the market
- the influence on fees would force some people to pursue courses other than their most preferred

Concerns expressed elsewhere in the consultation relating to scenario 3 can be summarised as:

- it is extremely difficult to predict what the skill sets we will need in the future and education tertiary education is not the only route to success
- some courses open a huge variety of career options - others are more aligned to a narrower professional identity (such as accountant)
- education isn't only to produce workers, it also produces citizens, creators, critics, and widen our collective knowledge and insight
- it is impossible to make predictions about the effect of a particular course on a particular student's job prospects

4.3.3.2 Constraints

Participants were asked: *What sort of safeguards, constraints, or changes would you add to the following scenario to mitigate your concerns*? Five participants contributed to this phase of the consultation and these contributions are shown in Appendix V, Table 18. Their contributions can be summarised as:

- make data available at both micro level to help inform individual choice
- make data available at macro level to assist decision makers to actively recruit for future needs
- make the model's predictions publicly available, but don't base course fees or admissions on the predictions. This lets individuals weigh the information provided with the model with other factors such as personal abilities and preferences
- make the integrated dataset available to researchers who can build a variety of different models to gain other types of insights

4.3.3.3 Benefits

Participants were asked: *Assuming appropriate controls were in place to mitigate your concerns, what do you consider to be the benefits of the following scenario?* Five participants contributed to this phase of the consultation. The benefits contributed are shown in Appendix V, Table 19 and can be summarised as:

- students would have a wider set of information available to them before making costly education decisions
- managing courses and course numbers is sensible

A comment from an earlier phase relating to this scenario can be summarised as:

- this may be useful in advising students when they are considering taking a course

4.3.4 Acceptability

Twelve of the 15 people that indicated they wished to be involved in the online consultation participated in the pre-consultation survey, though thirteen people participated in the first phase of the online consultation. Only six people participated in the post-consultation survey and the results in Table 6 below show how their levels of acceptability changed for each scenario.

As can be seen in Table 6, the mean and median level of acceptability for each scenario decreased following the online consultation. This can be attributed to participation in the concerns-focused phase of the online consultation being much greater than the subsequent phases that focussed on constraints to overcome the concerns, and the benefits of the scenarios.

Table 6: Changes in levels of acceptability for each scenario (where 1 = totally unacceptable and 7 = perfectly acceptable)

Participant	Scenario 1		Scenario 2		Scenario 3	
	Pre-consult	Post-consult	Pre-consult	Post-consult	Pre-consult	Post-consult
1	3	3	3	3	5	3
2	2	1	6	1	1	1
3	2	1	2	2	6	2
4	6	5	5	3	6	3
5	6	6	2	2	6	3
6	6	5	7	3	6	5
Mean	4.2	3.5	4.2	2.3	5.0	2.8
Median	4.5	4	4	2.5	6	3

4.4 Online consultation summary

The online consultation drew out the primary concerns of a group of experts, along with some suggestions for constraints and some acknowledgement of benefits associated with the scenarios tested. Overall, a lot of the core values around acceptability identified in public attitudes are also found in experts.

With lower levels of participation than desired, the full potential of the online consultation was not realised. After an initial peak in interest and engagement by the participants, there was insufficient sustained ongoing interaction during subsequent phases. This behaviour might be explained by a number of converging factors including short timeframes, the use of controversial scenarios, and the functionality of the online tool being too rigid to stimulate interaction and collaboration. However, the findings do provide some valuable insight into expert opinion in the matter of data integration.

In summary, acceptability thresholds for data integration include that:

- Integration acceptability is dependent on who can use the data, what for, and how the data is protected
- Data integration and use should be fair, accurate, representative and useful
- Data integration and use must represent public interests and values to avoid generating public mistrust and perverse outcomes
- Integrated data should be used in fair and meaningful ways
- Measures and indicators used should accurately represent reality and should not be overly simplistic

Suggested considerations for integrated data included that:

- Administrative data was not intended to be used statistically, so it may not be accurate enough and it will only capture people who interact with government departments
- It may not be appropriate or convenient to use data collected for one purpose for another
- A good quality integrated data system could be very expensive and time consuming
- Ensure the public are informed about intentions to integrate and use their data, and make sure these uses are widely acceptable

The participants' felt that if acceptability thresholds were met, and considerations were accounted for, then data integration could provide the following benefits:

- Better quality data and therefore, better quality information:
    - Longitudinal data would be available
    - Data would be more detailed and diverse
    - It could be more accurate than self-reported data (such as current census)
    - Data may be more current
    - Data could be updated more regularly
- Better information may then be used by researchers as well as government agencies to help with planning policies and services in a more effective way, which may also make them more efficient
- Better information may better reveal more about peoples' needs and vulnerable populations, which can then ensure these needs are then able to be met

Unfortunately, some of the ranking results suggest that participants were not fully engaged in the pairwise comparison process. The results indicate that the ranking exercise was unreliable for the following reasons: a small number of people participated in the pairwise comparisons (voting) for most phases of the consultation and some comments added near the end of the consultation phase were not involved in a sufficient number of pairwise comparisons to achieve a definitive ranking (these are indicated in each table of results); many of the contributions contained more than one idea, so it is not clear which idea participants voted for (this could have been overcome if participants actively cocreated the contributions to refine the ideas); and, in some cases the same idea was contributed more than once, yet the similar contributions did not achieve similar rankings.

The trend towards reduced acceptability for the scenarios following the consultation exercise is complicated by a low response rate so strong conclusions should not be drawn.

5 Discussion

5.1 Acceptability themes and thresholds

The multi-method approach adopted in this research allowed the researchers to encounter people from across New Zealand with a diverse range of opinions and life experiences. In general, participants' response to the research topic and process was highly considered, thoughtful, engaged and rational. In particular, the formats of the in-depth interviews and action research workshops were very successful in ensuring that the participants understood the processes around data integration. Participants were able to engage deeply with the topic and its practical implications, whether its application affected them directly or had implications for the wider population. Overall, the findings from the groups sampled here suggest that most New Zealanders have a relatively positive perception around data integration within public sector agencies (which is supported by recent New Zealand research, e.g. Lipps et al., 2010).

One particular area of interest for this research was to better understand how individuals from different demographic or social groups understand and react to data integration. Previous qualitative research has suggested less favourable attitudes amongst Māori, Pacific Islanders, and self-employed around perceptions of data integration (for example, Lipps et al., 2010). However, a key result from this study has been that attitudes to data integration do not appear to be strongly associated with particular types of social groups: there was wide variation of opinion within the particular interest groups, as well as across the full sample. While certain groups may better identify with the context of the specific examples raised, there was no evidence that the underlying level of acceptance was group-specific.

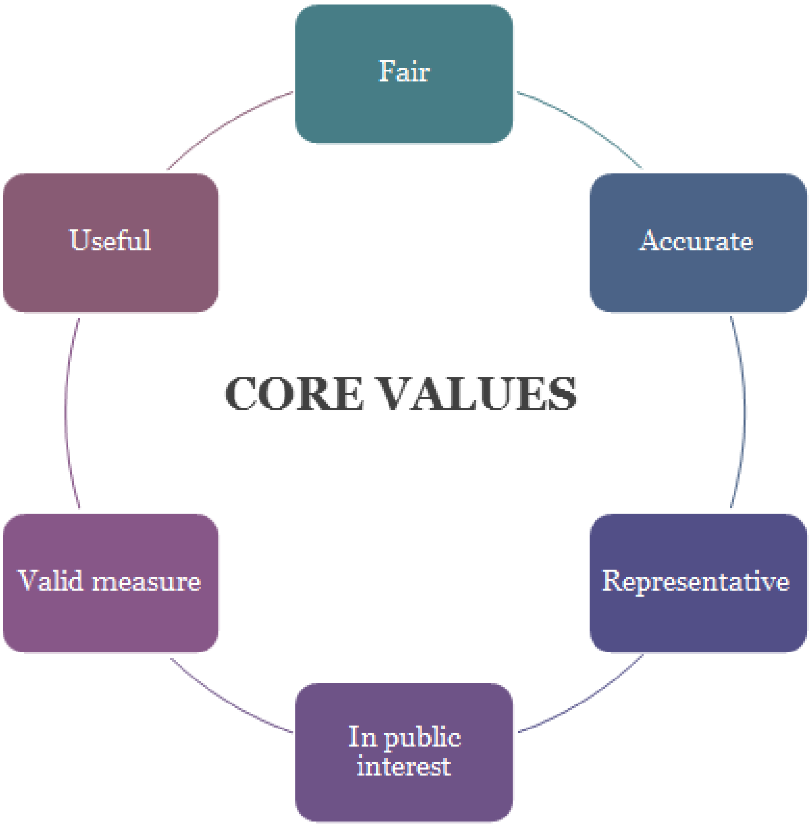
Acceptability of data integration appears to be largely influenced by the individual's own personal experiences with data sharing, typically through interactions with government departments and the type of outcomes that they have experienced. For example, many different types of people are concerned about the potential misuse of data to perpetuate negative stereotypes. Arguably those in more frequent direct contact with government agencies may have more examples or experiences to work from, which could explain the Lipps et al (2010) ethnicity and self-employment findings. Sherriff (2013) reports that the Office of National Statistics (ONS) of England and Wales similarly found that people's views around the use of administrative data was often based on personal experience, supplemented by negative media reports and that trust in the organisation is the most important factor affecting support for data integration.

Data integration acceptability was observed to exist along a continuum from totally acceptable to completely unacceptable. At one end of the spectrum, some participants had complete confidence in providing any and all of their personal information to certain organisations in good faith that their information would not be misused and were entirely comfortable with doing this. Others expressed considerable suspicion and questioned the need for and value of some or any of their data to be provided. Despite this wide variation, it is apparent that general data integration acceptability thresholds do exist.

Concerns were often based in personal experiences of being negatively stereotyped based on factors such as age, gender, ethnicity, occupation, criminal history, or where they lived. It was also often based on respondents' underlying values of data being used in a fair, beneficial, meaningful, appropriate and representative manner. These experiences and values did not appear to be specific to one type of person or another, but were found across the participants.

There also appears to be a more general value-based concern around appropriate use, as evidenced by participant concerns raised around data used to stereotype others outside of their own social or demographic group. Further evidence is that both expert and public attitudes align when it comes to core values. This could also mean that any positive messages or information intervention around data integration based on these values should work for both the general public and experts. Core values reveal a key part of understanding the underlying reasons behind why people shift along the continuum of acceptability.

Core values that underlie why data integration is acceptable



5.2 Acceptability shifts

During both the narrative interviews and the workshops there were examples where people's attitudes changed or were modified by additional information or context being introduced. In the case of the narrative interviews this was typically through the introduction of one of the interventions. For example, in the scenario around the re-located former criminal, when the scenario was first introduced a large number of participants initially stated that information about this person should be known by everyone living in the area. However, after adding the intervention that suggested the potentially negative outcome that data sharing could have for the former criminal, the participants frequently modified their position, to one where just the law enforcement agencies should be in receipt of the information. This represented a compromise that protected the residents' safety, as well as affording the former criminal a degree of privacy and the potential for a fresh start. The underlying value around personal and public safety remained the strongest acceptability criteria, but the values of fairness and privacy were also interacting at a secondary level.

Within the narrative interviews there were also examples of where there was a shift towards greater data sharing rather than restricted data sharing. One participant had narrated a very negative experience around wide-spread data-sharing following a court case that had had considerably negative outcomes for her personally. However, when their example was re-framed in the context of a statistical purpose where the information could be shared in a different way, for use by different organisations to inform and improve legal processes, they then approved of the wider sharing of information in this context.

Overall, people's attitudes tended to change toward being more supportive of data integration when they could see the benefits that would arise, or recognised the contribution of other factors that they valued, such as privacy and fairness. For example, in the workshops when they realised that statistical data integration could show how health outcomes were related to housing and inequality they felt that it would be good to integrate these types of data. This was particularly the case when the information needed is currently lacking, and the issue is topical in the media (due to recent deaths being attributed to cold damp social housing, overcrowding etc).

5.3 Factors that influence acceptability

The three key components of the information that appeared to be most influential for change in attitude were the type of information being shared, the purpose of sharing the data and the type of people / organisations who would be in receipt of the information.

Participants appeared to gauge acceptability primarily based on:

- *Why* the data needed to be integrated,
- *How* any personal or sensitive information would be protected,
- *What* the data would be used for, and
- *Who* would have access to and use the data
- Whether the benefits and value would outweigh the costs and risks

> *"I want to know who's holding my information, why they've got it, what they're going to do with it, and how secure is it"*

Data integration tended to be seen as relatively LESS acceptable if:

- There was no demonstrable need or purpose: the general sharing of all information was typically seen as neither necessary nor acceptable
- If it could be misused, or resulted in harmful or unfair outcomes, these include:
    - direct commercial gain
    - taking advantage of vulnerable people
    - profiling, stereotyping, or disadvantaging certain groups or types of people, or people from particular places
- Subjective, biased or limited data was used in a misrepresentative way
- It was used for spying and surveillance

> *"You wouldn't want a loan shark setting up in an area where there is a high level of poverty, people won't turn down a pot of gold, but it will be more detrimental to them in the long run"*

According to the 2012 Office of the Privacy Commissioner (OPC) survey, cited by Sheriff (2013), 40% of New Zealanders were concerned about the extent of Government surveillance of ordinary everyday citizens. Peoples' attitudes tended to change toward being less supportive when they could see the risks, including the risks of misuse of data (such as for commercial gain) or privacy/security/safety risks (particularly around being able to be identified individually, or having very personal data in a data base). Sherriff (2013) reports that ONS studies have also found that members of the public are primarily concerned about the possibility of their personal information be using against them through statistical applications of data integration, or through the sale of their information to commercial organisations.

Data integration tended to be seen as relatively MORE acceptable if:

- The need for and purpose of integration was specific and transparent
- It aligned with the greater public benefit and interests, and improved social outcomes
- The organisations involved were seen as highly trustworthy, professional, and competent
- If the data, metrics, integration process, and resulting uses were:
    - fair, appropriate, meaningful, accurate, representative, and of high quality and integrity
- Strict procedures and protocols exist around access and use
- Databases are completely secure and confidential
- Data is completely de-personalised and anonymous for statistical uses
- The public is informed about what is happening and why
- People are freely able to give and withdraw consent for their data being used, particularly if the data is personal or sensitive

Both the ONS (cited by Sherriff 2013) and Lips et al (2010) reported that acceptability of data integration was greater when there was a clear benefit to the public. When discussing the positive outcome 'for public good', participants described benefits such as things being done with the objective of making society better (fairer, easier, safer, improved service provision and promoting overall wellbeing). Efficiency improvements were generally seen as good, so long as they were also fair and economically viable, so that the cost of setting up systems did not negate the potential benefits to be delivered through their application.

Assuring anonymity by removing personal information, such as name and contact details, has been reported as an important acceptability criteria for data integration by a number of studies (Australian Bureau of Statistics 2010; ONS 2012; Scottish Government Social Research 2013, cited by Sherriff 2013).

While a key criticism of statistics and data use is that it can misrepresent or reinforce stereotypes, the corollary is that data integration and having access to accurate and meaningful data could help to dispel negative stereotypes, rather than reinforce them. For example, when stereotypes are inaccurate, using improved, integrated data to show that existing stereotypes are not a true representation of the actual fact.

> *"Everything's already stereotyped [...] so why not get it for real?"*

If data was integrated in an acceptable way, then participants generally felt that it could produce better operational and informational services.

Participants felt that acceptable operational and statistical data integration could result in:

- Better quality data and therefore, better quality information:
  - Longitudinal data would be more readily available
  - Data would likely be more detailed and diverse
  - More accurate and objective data
  - Data that may be more current
  - Data that could be updated more regularly

- Operational data integration could result in considerably more efficient and effective public services
- Better information generated from statistical data integration may be used by researchers as well as government agencies to help with planning policies and services in a more effective way, which may also make them more efficient
- Better information may reveal more about peoples' needs and vulnerable populations which can help ensure these needs are able to be met

> *"How far in the past does the sharing need to go? Some may be so old that it's no longer relevant. That shouldn't be shared if it's old, minor convictions."*
>
> *"I would be happy for the information to be used if it is going to help other parents... if it would benefit my child. I trust the government to look after the information.*

Reinforcing positive messages around data integration, the benefits of this, and stories about how better decisions or services have been provided in New Zealand as a consequence of data integration is very important. This is because it may take several positive messages to outweigh any negative perceptions or experiences, as negative impressions are typically quick to form and more difficult to dispel than positive ones (Baumeister, Bratslavsky, Finkenauer & Vohs, 2001). For example, it may only take one story in the media around problems with data privacy or confidentiality to influence trust in an organisation (e.g. Sherriff, 2013). Regaining trust is likely to take many positive experiences or examples, so proactively building positive stories can future-proof against a skewed negative view.

5.4 The influence of research techniques on findings

Narrative interviews

The narrative interviews extracted rich information around participants own experiences of operational data integration, the interviewer inquiry encouraged wider thought around the underlying issues, implications and alternative outcomes. Discussion around both operational and statistical examples allowed participants to select the agencies, data type, and purpose that did or could have contributed to the outcome, whether positive or negative, and provided the opportunity for participants to consider their criteria for an ideal solution, or acceptable outcome.

The scenarios used for the narrative interviews, both operational and statistical, were loosely worded to allow interpretation for either positive or negative outcomes. The participants could often see the potential for both depending on what information was used, how it was used and by whom. The environment created, where there were often no right or wrong answers encouraged variety and depth of responses and the opportunity to explore participants' attitudes from different perspectives. Participants own experiences fed very much into their conversations around acceptability, values and thresholds for data integration in typical day to day scenarios.

Workshops

The workshops employed coloured blocks to represent different types of personal information. Participants' had to physically move (share them) with Statistics NZ to demonstrate acceptability which created a less abstract environment for discussion around thresholds. The workshops effectively established which types of data people have more concerns around sharing, their fears around use of this type of data, perceived negative outcomes and their reasoning underlying these fears. The influence of participant's contributions and perspectives was clearly visualised through the placement and movement of blocks and reasoning 'talk' and explanation underlying the decisions. The workshops also allowed for individual contributions to influence the group discussion, where opinions of individual participants encouraged wider exploration of the issues by the group and the potential for participants to change their point of view based on the contributions and opinions of others.

Online Delphi

The online process, while it did not provide the interactive discussion it was designed for, did highlight some important lessons. Starting with controversial scenarios where the focus can be negative is difficult to then shift the focus to return to data integration benefits. Editing live scenarios or documents should be supplemented by separate live discussion or an ability to chat in a less formal way. In the context of a topic such as data integration, some form of facilitation in an online forum does seem appropriate to raise engagement and add prompts to push the conversation along when required. While an online approach does benefit from a potential wider pool of people, convenience (temporal and geographical), and reduces some biases through anonymity, it can have lower levels of engagement and interaction (Khodyakov, et al., 2011).

Overall the pros and cons of an online Delphi process need to be weighed against other methods. Depth of discussion and interaction in an online forum may be more difficult to capture with this process without some level of rapport built between participants. However, an online process is still beneficial for other aspects, such as to capture innovation and creative thinking. For example, capturing the potential wider uses and benefits of data integration, or even what data we should be capturing and why, and ranking the importance of this under relevant cost constraints.

5.5 Trust in Statistics New Zealand as a custodian of data

Most of the research participants had a very high level of trust in Statistics NZ as a professional, competent and trustworthy custodian of data. This finding supports work by Lips et al (2010), who also found that there was generally a high level of trust in Statistics NZ and associated government departments. Similarly, according to Sheriff (2013), there is likely to be more support for statistical uses of data integration when it is conducted by a statistical agency who also act as custodians of the data.

> *"Probably statistics is the only government department that can be depended upon to be completely independent and relevant."*

However, some of the research participants had very low levels of trust in Statistics NZ. Those with low levels of trust tended to feel this way if their experience with statistics was negative, such as if they felt that they personally, or where they live, had been stereotyped, stigmatised, or disadvantaged by generalisations based on statistics. Some also felt that statistics, in general, were not a useful or meaningful way to assess how well people are doing, or how happy they are, in a broader sense.

There was a varying degree of trust in other government agencies. Some participants voicing confidence in data integration by Government departments while others stated that they would find statistical data integration unacceptable if it was conducted by any organisation other than Statistics NZ. Participants cited the mishaps of other government agencies in handling personal data and high-tech systems as an example of 'what not to do'.

In order to retain and gain the public's trust, Statistics NZ should take every care to actively demonstrate that they meet the acceptability criteria held by the public, if, and when, they conduct data integration. Even reinforcing public understanding of basic principles, such as the data not being available for commercial purposes, and making the data security and anonymity processes clear and easy to find out about.

5.6 Data integration and the Census

While the Census was not the primary focus of the data integration findings, there were some highlights of relevance to the Census. Also, Sections 3.4.1 and 4.3.2 are of particular relevance. There was an overall perception that bringing in administrative data from other sources was seen positively by the public, in terms of accuracy, efficient government spending, and valuing the respondent's time.

> *"Well then you wouldn't need to fill out a census, and it would be more accurate than self-reported census data every five years, save time and money"*

The only element to consider is that some members of the public will be less comfortable with sharing types of information when they believe it is too personal, complex, or sensitive, or where they are not convinced that it needs to be shared, such as relationship status.

From an expert perspective there was a strong belief that the Census data could be supplemented by data from administrative sources (as opposed to replicated). There were benefits to having both approaches, rather than choosing either an actual Population Census or an "administrative data

Census”. Overall, parts of the Census could be shortened or removed to reduce cost and respondent burden, but these parts should be rigorously tested prior to removal.

> *“Use and continue to develop a combined IDI database to make good use of what is currently being collected, especially for research purposes, compare this with an actual census every 5 or at most 10 years. Re-evaluate this over time and refine if necessary the actual census questions to those that cannot be answered easily by other means. It will be many years (if at all) before NZ would be ready for solely an "administrative census".”*

Also, the theme around public wellbeing was raised as an important one around public engagement and improving the quality of the data and improving public response to the Census generally. Greater promotion of any wellbeing measures taken by the Census, or examples of how the Census supports studies that do lead to improved public wellbeing would likely improve engagement and responses.

> *“If the data they used was more meaningful and focused on peoples’ wellbeing then people may be more engaged in statistics and want to fill out the forms. If they ask nice questions, I want to give a nice real answer, and hopefully they’re going to use that to make things better.”*

6 Conclusions and recommendations

This research found that attitudes to data integration do not appear to be specifically associated with particular types of people, and that there seem to be consistent themes coming through around acceptable use. More fundamentally, data integration acceptability appears to be largely influenced by individual’s own personal experiences. Overall, if data is integrated following the processes outlined in the discussion and in line with the acceptability thresholds outlined below, then most participants felt that data integration is both useful and acceptable.

Summary infographic of core factors that influence acceptability

Less acceptable	⟷	*More acceptable*
Ambiguous & general	Purpose	Transparent & specific need
Insecure	Security	Secure
Open access	Access	Strict restrictions, procedures & protocols
Public/ personalised	Privacy	Confidential/ anonymous
Private	Interests	Public
Subjective & biased	Data & metrics	Representative & meaningful
Unfair & harmful	Outcomes	Fair & beneficial

Statistics NZ could consider the provision of positive messaging around data integration to ensure they retain and gain the trust of the New Zealand public, including:

- Messages of reassurance around continual improvement in privacy for operational data and confidentiality around statistical data. Even basic information reassuring people that the information met a clear and appropriate need, in particular around a public good as opposed to a commercial need.

- Demonstrating that the benefits would be greater than the costs and risks, and how risks could be mitigated (e.g. such as misrepresentative use of data, or use of data for surveillance).

- Where the data is personal, sensitive or complex in nature demonstrate how people could be offered the right to opt out.

- Examples highlighting the existing public benefits of data integration, including real examples of how the data has been successfully used and by whom, to improve levels of service for individuals or communities.

- Examples highlighting the potential public benefits if data is better integrated, including the opportunities to make more informed, fair, efficient and effective decisions to improve quality of life for New Zealanders. For example, the ability to use wider or less traditional metrics when determining how well people are doing or how happy they are.

6.1 Suggestions for further research and actions

Quantitative data could be collected to better gauge the strength and distribution of issues uncovered by the qualitative findings.

Investigating ways of effectively providing public reassurance and allaying people’s concerns around data integration.

Developing a publication that clearly demonstrates how those acceptability criteria identified as most important are addressed by Statistics NZ.

Developing a publication that clearly illustrates the different applications, uses and benefits of data integration.

7 References

Bauer, M. (1996). The narrative interview. *London School of Economics and Political Science Papers in Social Research Methods no. 1.*

Baumeister, R.F., Bratslavsky, E., Finkenauer, C. & Vohs, K.D. (2001). Bad is stronger than good. *Review of General Psychology, 5* (4), 323-370.

Bell, S., & Morse, S. (2010). Triple Task Method: Systemic, Reflective Action Research. *Syst Pract Action Res.*

Hsu, Chia-Chien & Sandford, Brian A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment Research & Evaluation*, 12(10). Available online: http://pareonline.net/getvn.asp?v=12&n=10

Khodyakov, D., Hempel, S., Rubenstein, L., Shekelle, P., Foy, R., Salem-Schatz, S., O'Neill, S., Danz1, M., & Dalal, S. (2011). Conducting Online Expert panels: a feasibility and experimental replicability study. *BMC Medical Research Methodology, 11,* 174.

Lips, M., Eppel, E., Cunningham, A. & Hopkins-Burns, V. (2010). *Public attitudes to the sharing of personal information in the course of online public service provision.* Available at: http://e-government.vuw.ac.nz/summary_IRD.aspx

Sagor, R. (2000). What is Action Research? In *Guiding School Improvement with Action Research.* ASCD.

Sheriff, R. (2013). *Privacy & Public Acceptability Research on the Use of Data about Individuals to Produce Statistics.* A review prepared for Statistics NZ. 28p.

Stein, M. K., & Bovalino, J. W. (2001). Manipulatives: One piece. *Mathematics Teaching in the Middle School*, 6, 356–359.

Vagias, W. M. (2006). *Likert-type scale response anchors.* Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.

Appendix I: Research Methods

The research comprised of three stages:

1. Thirty in-depth narrative interviews in Wellington and Christchurch;
2. Workshops with 30 participants across three urban centres (Wellington, Gisborne and Invercargill); and,
3. An on-line discussion to which 43 people were invited to participate.

The social groups of interest to Statistics New Zealand include: Māori, Pasifika, retirees, the unemployed, and the self-employed (Lips, Eppel, Cunningham & Hopkins-Burns, 2010). Recruitment for the interview and workshop stages of the research actively targeted these groups, but also included New Zealand residents who did not belong to these groups.

Each of the three research stages involved scenarios to facilitate meaningful discussion. The team at Statistics New Zealand developed the scenarios and in some instances, these were refined by the researchers.

In-depth Narrative Interviews

Current literature (Sheriff 2013) has established that it is difficult for non-experts to understand how and why data integration is necessary, and there is a tendency for people to default to 'knowns' when detailed explanations are provided. For example, when introduced to the details of data integration, it is common for people to default to an assumption that government agencies have always integrated administrative data. While this is a reasonable assumption, it does not represent an opinion from which the pros or cons of data integration can be delineated.

Question-response type interviews impose structures on the interviewee by selecting the theme and topics, driving the direction of the interview via the order of the questions asked, and by wording questions in the interviewer's language. In comparison, narrative inquiry reveals the perspective of the interviewee ("meaning making") using his or her own spontaneous language in the narration of events. As social groups tend to share a perspective of the world that is reflected in their lexicon, narrations can be assumed to preserve this perspective in a more genuine way. In addition, the ability to tell stories is relatively independent of education and language competence, which is particularly relevant to this research project where the social groups of interest vary from one another and from the social group that the interviewer belongs to.

As this research is seeking information from people who may not have clearly formed views about data re-use and integration, narrative inquiry provides a "safe" space for the interviewee to share any experiences they may personally have had with data integration and reveal their current understandings of and attitudes towards data integration.

The interview strategy outlined in Table 1 was trialed and refined with people from different population groups and follows the narrative interview format outlined by Bauer (1996), with the addition of an intervention stage to reveal thresholds of acceptability for data integration. The intervention activity focused on gaining more in-depth and quantifiable information about the acceptability thresholds that participants hold for a variety of hypothetical scenarios, both operational and statistical. It also provides an opportunity to identify what, if anything, can be done to mitigate participant's most serious concerns about the use of their data.

Table 1: Overview of In-depth Narrative Interview Process

Interview stage		Description of interview stage
1. Introduction		Orientation towards data collection and data integration through personal experience. Presentation of the central topic, including explanation of data integration, ruling out surveillance and spying.
2. Narration	2.1 Personal experience	Participant invited to share their own personal experience of data integration. Prompted to fill in gaps in the story.
	2.2 Other people’s experiences	Based on their positive/negative experience discussed, participant invited to share about another person’s experience with different outcomes.
	2.3 Statistical integration example	If the above stories only included operational examples: Real or hypothetical media story covering a statistical example of data integration presented to the participant and further narration encouraged to elicit their experience of similar stories.
3. Intervention	3.1 Scenario ranking activity	Participant presented with six scenarios on cards (three operational, three statistical) and asked to arrange them on a grid where one axis represents “acceptability” (with labels “acceptable” and “concerning” and the other “usefulness” (with labels of “useful” and “not useful”, with narration about how they came to choose each location on the grid.
	3.2 Increasing acceptability	Participant asked to consider the least acceptable scenario and what, in an ideal world, would make it more acceptable. Includes prompts for organizational responsibility, assurances, and standards.
4. Close		Small talk to summarize and reflect. Includes invitation to receive study summary, and SNZ materials about data integration.

Participants were presented with cards representing six different scenarios (three operational and three statistical scenarios) and asked to rank the scenarios on two axes, according to (a) how useful or not useful they consider the outcomes to be and (b) how acceptable or concerning they consider the outcomes to be. During the process of ranking the scenarios, participants were invited to provide a narrative around their decision-making and choices. The process of ranking the scenarios and the associated narratives provided a graded assessment of acceptability thresholds through the range of concerns and benefits that each participant associates with specific scenario. Supplementary information was then provided through the presentation of a number of interventions with different possible outcomes for each of the scenarios. Participants were asked to discuss any changes in their ranking of the scenario in term of the two axes, usefulness and acceptability, in the light of the new information. This provided the opportunity for the participants to explore the scenarios from different perspectives where sometimes there might be a perceived benefit and at other times a perceived harmful outcome, and where sometimes there might be a benefit for the individual but a cost for the wider public. Where a particular scenario and/or intervention was graded by the participant as concerning and not useful participants were encouraged to explore what additional information or intervention is needed in order to mitigate their concerns and make the scenarios more acceptable. The scenarios and interventions are provided in Appendix II.

Thirty interviews were conducted across Wellington and Christchurch. Participants were recruited using a convenience sample approach where not-for-profit organisations such as schools, charitable trusts, community groups and sports clubs whose membership was likely to include people belonging to the social groups of interest were contacted for assistance in identifying people interested in taking part in the study. In recognition of their assistance in recruiting participants, the organisation was offered a donation towards their fundraising and this incentive could be split between the participant and organisation. This method has successfully been used in the past to recruit participants from hard to reach populations (Ivory V, Burton J and Harding A. 2013. *Testing Successful Central City Living in Christchurch: What will it take for people to live there?* Opus International Consultants, Wellington). The narrative interviews took 45-60 minutes to complete.

Workshops

The initial intention for this stage was to adopt an action research approach in which participants are asked to join and separate data fields. In addition, it was proposed that participants be intercepted outside locations where people belonging to social groups of interest were likely to meet. Following discussions with Statistics New Zealand, and in consideration of the findings from the narrative interview and online consultation stages, a workshop approach and targeted recruitment was adopted instead. In particular, Statistics New Zealand wanted to:

- Identify information that population groups of interest may be sensitive about sharing with Statistics New Zealand and understand the reasons for this;
- Explore thresholds of acceptability for potential Stats New Zealand activities, including an administrative Census;
- Ensure that participation on behalf of the researchers would be restricted to neutral facilitation rather than active collaboration between the researchers and participants;
- Ensure participants are explicitly aware that the topic is the integration of individual records, and not merely the comparison of aggregate data;
- Place emphasis on the individual nature of data linking; and,
- Capture dynamic interaction with data, as opposed to a conversation or conceptual interaction.

As action research methods require a collaborative research approach in which both the participants and researchers contribute to further learning and understanding for both parties (Bell & Morse, 2010), a researcher-neutral workshop methodology was embraced. Direct recruitment of workshop participants was pursued as it is a more reliable approach than interception for the following reasons:

- It can be difficult to successfully intercept participants when the topic is complex and difficult to understand, especially given some of the social groups of interest. A benefit of direct recruitment is a higher likelihood of capturing these groups;
- The length of time required to undertake a walkthrough is prohibitive with regards to intercepting the public (typically intercepts are about 5 minutes long; 25 minutes for people already very engaged in the topic);

- Better quality data as participants have committed to a time slot, so the results are less likely to be biased by people rushing to finish; and,
- As the workshop is a group process, it will capture some of the interaction that could not be captured remotely in the online consultation stage of the research.

The workshop stage involved different and independent participants to the in-depth narrative interviews.

As a topic for discussion, the complexity of data integration is high. The workshop method was developed with the intention of un-teasing complexity by direct observation of participant reaction to data integration using concrete manipulatives (children's building blocks) to represent data fields. Concrete manipulatives can effectively meeting varying learning styles and facilitate concept exploration in a group setting (Stein and Bovalino, 2010). Concrete manipulatives enable participants to link information provided during the workshop with personal opinions and knowledge in a meaningful way. The plastic blocks were identical in size and shape and coloured coded to represent the government agency for which the data was originally collected, with the data field written on each block. The data fields were personalised (see Appendix III), for example "my relationship status" rather than "relationship status" to emphasis the linking of data about each participant.

Each workshop was facilitated by a lead researcher, while a silent observer recorded observations and photographed the outcomes of each block sorting exercise. Participants could ask the researchers for clarification about what the wording on a particular block represents. However, the researchers were careful not to lead the participants in any way with positive or negative information (i.e. no contribution about how a data field could be combined with another for a certain purpose).

Four workshops were conducted, for which 30 participants were recruited (oversampling ensured that the minimum target of 20 participants was achieved). The intention was to conduct workshops in Auckland, Gisborne, Wellington, and Invercargill. Unfortunately, recruitment of Pasifika people in Auckland proved difficult given the research timeframe, so the fourth workshop was conducted in Wellington instead. This spread in location provided a mix of participants from the target populations from urban centres that vary in size and geographic region. Recruitment for each workshop focused on three of the six population groups of interest, recognizing that some participants will represent more than one group. In particular, Māori and Pasifika participants were involved in separate workshops as the findings of the in-depth narrative interviews indicated different perspectives and greater ease sharing amongst people of their own ethnicity. Capturing population groups in one large urban location and one smaller urban can: (a) facilitate good quality discussion amongst diverse groups, (b) allow comparisons between different population groups, and (c) allow comparisons between large cities and smaller regional centres. In contrast, hosting a homogenous group from a single location may not capture the diversity of opinion that exists across the country.

Workshop process

An overview of the workshop process is given in Table 2.

The introductory information provided was brief:

> *Data integration happens when your personal information held by different government agencies is shared with Statistics New Zealand. All the information about you is then linked*

> *up. Statistics New Zealand would not pass on your individual identifying information to other government agencies or users though. Integrated data will help to identify population trends, like the number of jobs people have over their lifetime.*

Pieces of named, laminated paper were used to indicate the agency of origin for each data field (i.e. the agency that originally collected the data). The Statistics New Zealand space was a larger piece of paper in anticipation of potentially receiving a large number of blocks. Initially, the blocks were grouped with their agency of origin, so the blocks already collected by Statistics New Zealand via the Population Census started out in the Statistics New Zealand space.

The first block sorting activity provided baseline insight into the types of data fields the participants were comfortable sharing with Statistics New Zealand. Participants were invited to move blocks that they are willing to share to the Statistics New Zealand box, to leave them with the agency of origin to indicate an unwillingness to share the data field, or to move blocks to a “not sure” box. The facilitator then proceeded to check that every participant was comfortable with the location of each block, or if the block should be moved to a different space. This was done systematically, first with the blocks in the Statistics New Zealand space, then the blocks left with their agency of origin (the “no share” blocks), and finally the “not sure” blocks. The facilitator was careful not to provide any information, rather to prompt discussion by asking neutral questions only.

Participants were then immersed in realistic scenarios, allowing them to further explore how and why they would permit data integration. The use of valid and reliable scenarios helps participants to become engaged with the research process and aligns any findings with research outcomes (Sagor, 2000). Whilst in the in-depth narrative interview stage the scenarios were a combination of statistical and operational uses of data, the workshop stage focussed on realistic statistical activities that Statistics New Zealand might perform. Six scenarios were developed, but only as many scenarios as the workshop timing allowed for were used. The scenarios were presented in the same order each time. Before commencing the first scenario, the “no share” and “not sure” blocks were marked with sticky dots.

For each scenario, only the relevant blocks were used (refer to Appendix IV). The participants were presented with the group of blocks required for each scenario and asked what they think the advantages and disadvantages of each scenario are, and if this would change their willingness to share any of the blocks with Statistics New Zealand. This process was repeated with further scenarios, ensuring that the workshop did not progress for more than an hour in duration.

Table 2: Overview of Workshop Process

Workshop stage		Description of workshop stage
1. Introduction		Brief overview of what data integration involves.
2. Main block sorting	2.1 Initial sort	Participants were asked to identify the information they are comfortable sharing with Statistics New Zealand (SNZ) by physically moving such blocks representative of personal data fields from the agency of origin to SNZ. Any blocks participants definitely don't want to share with SNZ are left with the agency of origin. Any blocks participants aren't sure about are moved to the "not sure" space. Discussion encouraged.
	2.2 Check "shared" blocks	For each block in the SNZ space, participants were asked: 1. if everyone is comfortable sharing it with SNZ. 2. if they can think of any circumstances where they would not be comfortable sharing it. 3. if they want to leave it or move it back to the agency of origin or "not sure" space
	2.3 Check "no share" blocks	For any blocks that remain with the agency of origin, participants were asked: 1. why they don't want that information shared with SNZ 2. if they can see any reasons it might be useful for SNZ to have access to that information 3. if they want to leave it or move it to the SNZ or "not sure" spaces
	2.4 Check "not sure" blocks	For any blocks in the "not sure" space, participants were asked: 1. why they are uncomfortable sharing that information with SNZ 2. if they can see any reasons it might be useful for SNZ to have access to that information 3. if they want to leave it in the "not sure" space or move it.
3. Scenarios		For each of up to six statistical scenarios (depending on the time available), only the relevant blocks for each scenario are used. Participants are asked what they think about the scenario, prompting for discussion about advantages and disadvantages in a non-leading manner. For any blocks that participants had previously been unwilling or "not sure" about sharing, asked if they would be willing to share this will SNZ in the context of the scenario, or if there are any blocks they are no longer willing to share in the context of the scenario.
4. Close		Includes invitation to receive study summary, and SNZ materials about data integration.

Online Consultation with Experts

While interviews with individuals were intended to reveal a depth of information about how people understand data re-use and integration (or react to knowledge of data integration), it is also important to determine why and where experts harbour concerns about data integration.

The online consultation phase adopted a modified Delphi approach to explore concerns, constraints to address the concerns, and benefits associated with a selection of controversial data integration scenarios. The Delphi Technique is a method of structured communication in a workshop setting in which a panel of experts answer questions in two or more rounds. First the questions are answered individually, then discussion and further rounds lead to a group consensus (Hsu & Sandford, 2007).

The online method was conducted using an ideas generation package called Codigital, which provides a platform for real-time iterative group engagement. The Codigital process asks individuals to submit ideas in relation to a question or topic. As ideas are submitted, individual participants are presented with pairwise comparisons and vote to rank the group's concerns. Any individual can suggest edits for the articulation of an idea, and these edits are also voted on by the participants. Both processes work in parallel, and once edits are complete the new idea is ranked with the rest via pairwise comparisons. This process offers a type of structured group conversation in which participants remain anonymous as they articulate and rank concerns. There are several benefits to this process:

- Anonymity naturally encourages participation and reduces bias associated with dominant personalities when an exercise of this nature is conducted in-person;
- This process allows participants to revise earlier responses in light of others perceptions; and,
- A form of consensus is gained amongst experts on priority areas of concern.

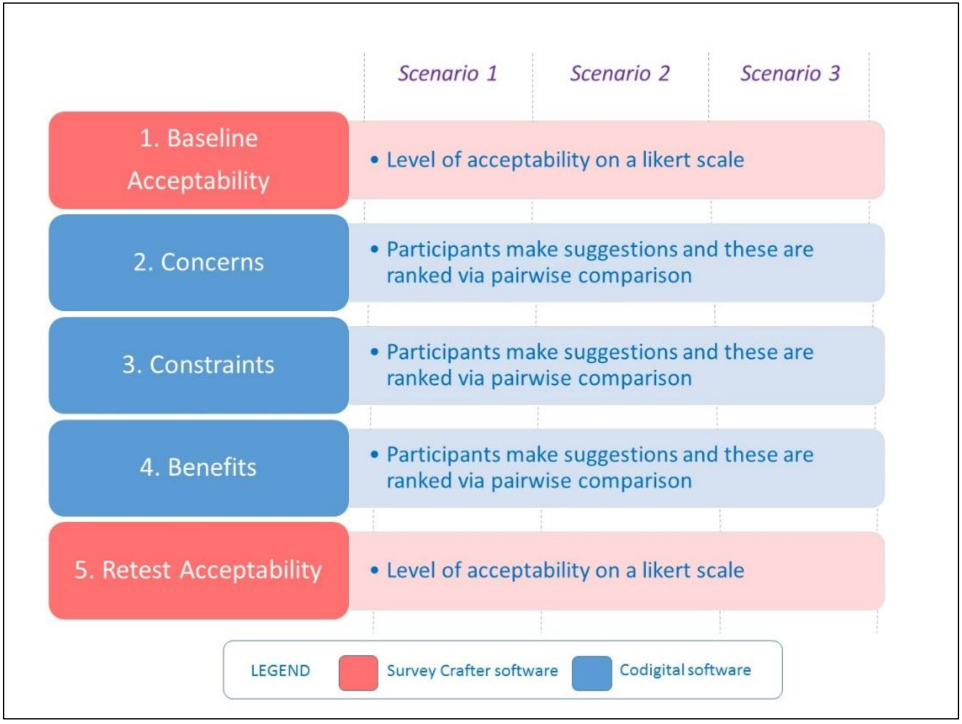
The online consultation process is presented schematically in Figure 1 below.

Levels of acceptability for each scenario were captured before and after the online consultation in an online questionnaire programme called Survey Crafter, using the likert scale developed by Vagias (2006) where 1 = totally unacceptable and 7 = perfectly acceptable. The initial levels of acceptability would provide a baseline understanding of each scenario. Controversial data integration scenarios (two statistical and one operational) were used to gauge how new information generated through the discussion process influenced each participant's acceptability of the scenarios.

Participants were then presented the same scenarios in Codigital. First, participants were invited to submit their concerns about each scenario. When that exercise was complete, they were asked to submit ideas about how the most highly ranked concern/s could be mitigated for each scenario. Finally, they were asked to contribute their ideas about the potential benefits of each scenario. Each of these three consultation phases occurred over a 46 hour period (8am on day 1 to 6pm on day 2, allowing for cogitation time overnight). The first phase was run on a Thursday and Friday, and the subsequent phases Tuesday-Friday and Thursday-Friday during the following week. The Monday was a public holiday, and the second phase was kept open to contributions during the third phase because the response rate during the intended Tuesday-Wednesday period was low.

Once the contribution periods in Codigital were complete, the participants were redirected to Survey Crafter and again asked to indicate their level of acceptability for each scenario.

Figure 1: Overview of the Online Consultation Process



The three controversial, hypothetical scenarios were as follows:

a. *A single view of citizen costs*
What if we could use a large integrated dataset to measure how much each person is costing their fellow citizens/taxpayers? We could use information from agencies such as Ministry of Social Development, Justice and IRD to track and/or predict the cost of services over time. We could also see if people use more government support than their taxes cover.

b. *An alternative census*
What if we replaced Census data with data from administrative sources? You wouldn't need to fill in a Census form anymore. We would know all about you, all of the time, by using information from your interactions with government departments like Inland Revenue, MSD, or the Ministry of Health. We would use this information to create Official Statistics on a monthly, quarterly or annual basis instead of every five years.

c. *Targeted education*
What if we used an integrated dataset to predict the likelihood that you'd be able to get a job in your chosen field? The statistics produced would mean we could then charge variable costs for education to manage demand for courses, or to limit course numbers.

It was evident during the first phase of the consultation that participants were contributing multiple ideas in one comment, complicating the ranking process. In subsequent emails communicating access to the next phase of the research, participants were reminded to submit each novel suggestion as an individual idea. One idea contributed to Scenario 1 in the first phase was split into two ideas by moderators. However, participants continued to contribute multiple ideas per comment.

The participants also proved reluctant to cogenerate (edit and refine) the ideas contributed, so the interactive potential of the software was not realised. Participants were also reminded in subsequent communications that The Codigital process is designed to be iterative and interactive and to edit or add to existing ideas to improve them. This failed to encourage cogeneration of ideas.

Appendix II: Scenarios for Narrative Interviews

Table 7: Operational scenarios and Interventions for narrative interviews

Operational Examples and interventions	
1. Recently widowed – data sharing can be used to make transition easier, for instance information agencies like IRD that a person is deceased. This will mean grieving relatives aren't pestered with insensitive or intrusive correspondence	+ Sharing news of his death could help reduce the amount of paperwork required to be completed by his wife. + Sharing news of his death could help prevent his wife from receiving intrusive or insensitive correspondence. — Sharing news of his death might lead to insensitive correspondence about changes in the wife's entitlements or allowances. — Sharing news of his death might lead to intrusive enquiries about his widows' ability to care for / support herself
2. Personal and community security – sometimes people who hurt others are known to authorities before that person acts. Should action be taken to prevent harm?	+ Sharing information about this person's whereabouts and criminal history could encourage vigilance and help keep the community safe. + Sharing information about this person could enable the police to more closely monitor his/her activities / behaviour and prevent an incident — Sharing this information could prevent the person from having a fresh start within the community. — Sharing this information could make it difficult for this person to find a job.
3. Children at risk of underachieving in education – it could be possible to identify kids who are at risk of not completing education	+ Sharing information could reveal an underlying reason (e.g. illness or caring for a family member) for the drop in the performance and attendance that can be addressed.

	+ Sharing information could help establish appropriate support for the child, parents and teachers to maximise the child’s learning potential. + Sharing information could show that the child has been to many schools in the past two years and will need extra support to catch up. — Sharing information could put the child at risk of abuse at home. — Sharing information could prompt the parents to move away causing further disruption to the child’s education.

Table 8: Statistical scenarios and interventions for narrative interviews

Statistical Examples and Interventions	
1. How’s my suburb? - Data can be pulled together to profile suburban areas, including housing, social and population data	+ The local council could use this information to prioritise spending on the services identified as most needed. + Small businesses could use this information to identify potential markets for their services — Real estate agents could use this information to boost prices in particular areas and discourage sales in others.
2. Census mortality study – Linking Census to health data can show how individual lifestyle and environment contributes to illness	+ The Ministry of Health could use this knowledge to plan a health promotion campaign or prioritise spending. - Insurance providers could use this knowledge to raise premiums or deny cover. - Doctors could use the knowledge to ration expensive treatments.

3. How are my people doing? – data can be integrated to provide a profile of how people like you are faring, e.g. older people, Māori, Pasifika, Iwi, women	+ The Council could use this information to plan services. + Local businesses could use this information to help target their services for customers – Loan sharks could use this information to target groups within the community

Appendix III: Personalised Data Blocks

Table 9: Personalised data blocks for workshops

Agency of origin	Personal data field blocks
Ministry of Health	1. Number of and reason for visits to Hospital or GP 2. Medical Conditions I've had 3. My ethnicity 4. Whether I'm a smoker (or drinker) 5. My address history
Inland Revenue Department	1. My exact income every year 2. My address history 3. If pay or owe any child support
Ministry of Social Development	1. If I've ever been on a benefit 2. How many and the type of allowances I've claimed 3. My relationship history and status 4. My address history
Statistics New Zealand (collected as part of the Population Census)	1. My income 2. My religion 3. My household type and number of people in my home 4. My occupation 5. My ethnicity 6. My relationship status (i.e. married, divorced)
Department of Justice	1. My history of convictions 2. My history of interaction with Police 3. My sentence history
Department of Internal Affairs	1. My birth date 2. My family's death records 3. My family's birth records 4. My marriage records
Ministry of Education	1. My educational history 2. The amount and repayment progress of my student loan
Ministry of Business Immigration and Employment	3. My migration status 4. How often and when I enter and/or leave NZ

Appendix IV: Scenarios for Workshops

Table 10: Scenarios and relevant blocks for workshops

Scenario	**Block**
1. Statistics NZ could use integrated data to count and describe the NZ population like they do with the Population Census, but you wouldn't have to fill out a form on Census night anymore.	How many and the type of allowances I've claimed If I've ever been on a benefit My address history My birth date My educational history My ethnicity My exact income every year My migration status My relationship history and status Whether I'm a smoker (or drinker)
2. Integrated data could be used to identify neighbourhoods that might benefit from more community policing	If I've ever been on a benefit My address history My educational history My ethnicity My history of convictions My history of interaction with Police My household type and number of people in my home My occupation My relationship history and status My sentence history Number of and reason for visits to Hospital or GP
3. Statistics NZ could use your Inland Revenue number or NHI number to get more accurate details to add to their consults.	Medical Conditions I've had My address history My ethnicity My exact income every year Number of and reason for visits to Hospital or GP Whether I'm a smoker (or drinker)

4. A private research company could use integrated data from public sources to study health trends in different industries (for example, comparing the health of office workers and forestry workers). *If participants speculate about why a private company would study this. If so, researchers suggest the findings could be published for public good or to inform commercial use.*	If I've ever been on a benefit Medical Conditions I've had My address history My birth date My educational history My ethnicity My family's birth records My family's death records My history of interaction with Police My household type and number of people in my home My income My marriage records My occupation My religion My sentence history Number of and reason for visits to Hospital or GP Whether I'm a smoker (or drinker)
5. Integrated data could be used to produce statistics on long-term outcomes for at-risk youth	My household type and number of people in my home Medical Conditions I've had My birth date My educational history My ethnicity My family's birth records My family's death records My history of convictions My history of interaction with Police My sentence history Number of and reason for visits to Hospital or GP Whether I'm a smoker (or drinker)

6. Integrated data could be used to identify neighbourhoods that might benefit from social housing or social services	How many and the type of allowances I've claimed If I've ever been on a benefit If pay or owe any child support Medical Conditions I've had My address history My birth date My educational history My ethnicity My exact income every year My family's birth records My family's death records My history of convictions My history of interaction with Police My sentence history Number of and reason for visits to Hospital or GP

Appendix V: Online Consultation Outputs

How to interpret the results tables

The results tables in this section display the **rank** of each **idea** contributed, which is an indication of importance placed on the idea by the group of experts. The ideas are presented verbatim and interpreted in the text accompanying the tables. **Cocreating count** represents the number of participants who collaborated on the wording of the idea (both changing content and voting to accept or reject the content changes). The participants were generally reluctant to cocreate the ideas and in some cases cocreation was absent. Where this occurred, a cocreation column is not included in the table. Finally, **No. of voters** represents the number of participants who collaborated on the ranking of the idea (voting).

Scenario 1

Scenario: *What if we could use a large integrated dataset to measure how much each person is costing their fellow citizens/taxpayers? We could use information from agencies such as Ministry of Social Development, Justice and IRD to track and/or predict the cost of services over time. We could also see if people use more government support than their taxes cover.*

Table 11: Concerns contributed for Scenario 1

Rank	Idea	Cocreating count	No. of voters
1	Measuring cost and usage of Government services - when done at a class level and through ethical antalytics is fine; it is when this is used to profile and/or stigmatise individuals or clasees of people that I think we have crossed a line. the addition of the "We could also see if people use more government support than their taxes cover" component of the scenario moves into this more inappropriate use of data.	3	9
2	Elderly people, sick people, students, children, voluntary workers, fulltime mothers, etc. would appear as burdens on society. This idea assumes a person's contribution to NZ inc. can be accessed financially. It also assumes everyone must be "producing" at all times to be "of worth" to society.	5	11
3	What useful action could we take on the basis of categorising citizens into "tax positive" vs "tax negative"? None that I can see. The very framing of "getting more than you give" around Justice and Social Development is broken. The costs of *not* funding Justice and Social Development exceed the costs of doing so. We are not a set of individual contributors or beneficiaries to a single pool of money, we are people whose lives are fundamentally intertwined: as neighbours, customers, employers, citizens, and many more roles. To boil that down to an economic transaction is to invite the optimisation of that transaction for the benefit of those who are found to be paying more tax than they are receiving direct benefit: aka, those who are already rich in wealth and opportunity.	1	8

4	The bias isn't necessarily about class, but about what's quantifiable. For example, indirect taxes just can't be tracked with administrative data, and dropping $15b of tax revenue leads to some pretty skewed conclusions: http://polity.co.nz/content/key-pathalogical-liar-about-tax It's particularly misleading when it's framed as a "complete" view.	1	9
5	This feeds into the idea that "my taxes pay for beneficiaries" and I guess superanuitants. To review costs on this basis would mean data would need to be collected on a persons contribution to tax as well as a persons cost . It indicates that there is a one to one relationship	4	10
6	this creates a scientifically justified underclass	1	11
7*	One of the greatest risks posed by concepts of linking and centralising social data, is that there can be a 'single source of truth'. This is a childish concept. Data needs competition and evolution just like anything else. People must constantly strive for better understanding, and better ways to gain insight. Many of our greatest failures have been based on flawed models. So the premise of many of these thought-experiments is fundamentally flawed. A centrally managed data model based on some sort of super-architect would cause significant strategic damage to our competitiveness. The real answer is to get agencies to use the data they have. In the long term centralised "approved" architectures simply give the illusion of understanding, and promote lazy thinking.	1	2
8*	Analysing cost data is very complex and no matter how many datasets are in the IDI, there would not be the data available that would measure a person's costs and benefits to society, as not everything is monetary (thank goodness!)	1	0

* An idea that was recently submitted and has yet to receive a sufficient number of votes to be properly ranked

Table 12: Constraints contributed for Scenario 1

Rank	Idea	No. of voters
1	Instead of calculating fiscal cost at the individual level, develop broad indicators of future fiscal cost. For example, interacting with the youth justice system could be an indicator of future fiscal cost. This will give government departments useful information about where interventions could be targeted, without shaming individuals.	3
2	Remove the comparison of government support with taxes paid. Reasons: 1. I don't see how this information would help government departments target interventions, 2. It would be incorrectly viewed as a person's cost to society, leading to social tension, 3. the purpose of a social safety net is to redistribute resources to those in need.	3
3	Do not release this information publicly. The benefit from this exercise is government departments (or the state sector as a whole) being able to target interventions at people that are likely to require government support in the future. Releasing this information publicly is not necessary for this, and would mostly just lead to shaming of people in need (i.e. beneficiary bashing).	3
4*	With cost as the primary determinant there is a risk we lose the wider public value of investing in public services	2

	I would suggest that we balance the 'cost' factor with other qualitative measures of public value. (i.e it costs us $yM to deliver x service, in return we expect less/more positive social outcomes (eg; less violent crime, less truancy, more succesful education outcomes).	

* An idea that was recently submitted and has yet to receive a sufficient number of votes to be properly ranked

Table 13: Benefits contributed for Scenario 1

Rank	Idea	No. of voters
1	I simply can't see anything useful that would come out of such a proposition.	4
2	This would be useful for broadening the view of policy analysts beyond their own client admin data analysis. If someone is 'off-benefit' for instance it would be tempting to assume they are better off. By extending the view to cover employment information such as income better assessments could be made about transitional requirements for beneficiaries returning to/entering work.	4
3*	This would be an excellent tool with which to identify and punish 'undeserving' groups of people.	2
4*	We have a large database which is partially populated with Monte Carlo values. The data are therefore only accurate when aggregated over e.g. meshblocks. Despite this, there is strong demand and there have even been circuitous attempts, to retrieve individual information points. We found that once individual data exists, users will demand this data even when the data are not permitted to be used in a non aggregated way.	1
5*	This can be useful for: 1. identifying vulnerable population groups, 2. finding early indicators of future use of government services, 3. looking for areas where early interventions could lead to future fiscal savings.	2
6*	This could help 'defend' programs that target vulnerable populations by illustrating their potential benefits. "This program costs $20 million per year, but based on our modelling the target population could cost the government $500 million in the long run. The small initial benefits can build to large fiscal savings in the long run."	2

* An idea that was recently submitted and has yet to receive a sufficient number of votes to be properly ranked

Scenario 2

Scenario: *What if we replaced Census data with data from administrative sources? You wouldn't need to fill in a Census form anymore. We would know all about you, all of the time, by using information from your interactions with government departments like Inland Revenue, MSD, or the Ministry of Health. We would use this information to create Official Statistics on a monthly, quarterly or annual basis instead of every five years.*

Table 14: Concerns contributed for Scenario 2

Rank	Idea	No. of voters
1	To create an accurate census from administrative data, especially if you want to allow the random setting of a census date, a lot of effort is required to constantly keep track of people to obtain the	7

	correct denominator - who is in the country, where they are living, with whom etc. The IDI has the potential of being very powerful and useful for research analyses, but also very frightening if it is used for "Big Brother" purposes. It cannot fulfill all the roles of a standard 5 or 10 year census, especially as a lot of the valuable information will not be recorded on any other government databases.	
2	In the absence of a unique individual identifier that follows individuals across systems (which is unlikely to be publicly acceptable), relying on probabilistic linkages across the various IDI admin datasets to replace the census is extremely risky. The census will become a byproduct of admin data collected for various reasons, and managed by agencies which may have poor institutional co-operation - SNZ will have little power to control those data collection practices. In NZ we rely heavily on ethnicity data. To my knowledge no country in the world that has shifted from a enumeration census to an admin/register model collects ethnicity data. This is a black hole that is poorly understood globally but very important for NZ. These issues aside, I think both SNZ and the govt have misjudged the 1acceptability of unit-level data linkage to the NZ public. It has an under-t1he-radar feel about it and could seriously damage trust in the OSS	7
3	As I understand it, there is valuable information collected by the census which can not be found (or is not as accurate) in administrative data sources; e.g. ethnicity, religion, family structure	9
4	data provided for one purpose (i.e getting a drivers licence, passport etc) is given within the context of that service being demanded. There is a significant jump in logic (or the collection processes for those other disparate services need to be carefully (re)constructed) to then take that data and use it for another purpose.	9
5	What data protections would be in place to allow for census-related data to be pulled cleanly, and would those who have little interaction be underrepresented in this data?	8
6	I can definitely see why it'd be useful, but I'm getting super creeped out. The IDI was supposed to be an anonymised collection of unit-level records which could be aggregated as required in strictly controlled and monitored conditions. This sounds different, like the unit-level records across government is readily accessible. I know that the difference between the two is trivial - linking unit-level records across government is a fundamental part of the IDI, and so by definition any IDI would allow linked unit-level records to be extracted. But how accessible that linked unit-level record is, who can access it and under what conditions are really important to me.	9
7	People who have no interactions with government departments would be missed.	10
8	The above comments capture well my concerns. A census at a fixed point in time will give a far more accurate picture of our demographics. It may well be worth doing as per the scenario but also have the regular periodic census with the scenario complementing it.	7
9	I am glad you are allowing discussion of these scenarios as you may agree/disagree with part of the scenario but strongly disagree/agree with a later sentence. Teasing out the underlying thoughts is important.	7

Table 15: Constraints contributed for Scenario 2

Rank	Idea	No. of voters
1	Use and continue to develop a combined IDI database to make good use of what is currently being collected, especially for research purposes, compare this with an actual census every 5 or at most 10 years. Re-evaluate this over time and refine if necessary the actual census questions to those that can not be answered easily by other means. It will be many years (if at all) before NZ would be ready for solely an "administrative census".	3
2	Keep working on IDI as the frame for an administrative census but test it rigorously against the five-yearly census, using at least 2 time points. This testing phase is vital to ensure that it is both fit for purpose and that there is broad pubic awareness and acceptance of how administrative data are being used. Don't underestimate public	3

	acceptability or the time it takes to get the right systems in place. The shift to an administrative census has taken decades in most countries; Austria is the only country that has shifted from a footwork census to an admin model in a decade and the investment was huge.	
3	Build the "administrative data census" and compare it to the actual census. In areas where the two line up well (preferably comparing to more than one census), those questions in the census could be shortened or removed to reduce cost and respondent burden.	3
4	there are issues with both models . Not everyone fills in the census form and people do not necessarily give adequate data . It would be useful to have updated data from other sources eg health on a more regular basis to give input onto strategic planning	4

Table 16: Benefits contributed for Scenario 2

Rank	Idea	No. of voters
1	Research could be done on the linked administrative databases, creating insights of peoples lives and their various interactions (but we still need a standard census conducted as well).	4
2	The data would be longitudinal.	4
3	Benefit 1: Detailed population data at a geographical level sub-five yearly. Benefit 2: Potential for service population information	4
4	So the assumption is that data would be available from transactional sources to provide a 'rolling census'.	5
5*	If it was a genuinely comprehensive, high-quality 'real time' linked dataset that could robustly match people to their main residence, the possibilities for research, policy and interventions would be endless. I don't see how it could work as a census substitute without a central register of dwellings.	2
6*	If data were tracked from all sources it would probably be more comprehensive than the census . It does assume that everyone is linked into some government data . Housing data is not well collected at this stage	2

* An idea that was recently submitted and has yet to receive a sufficient number of votes to be properly ranked

Scenario 3

Scenario: *What if we used an integrated dataset to predict the likelihood that you'd be able to get a job in your chosen field? The statistics produced would mean we could then charge variable costs for education to manage demand for courses, or to limit course numbers.*

Table 17: Concerns contributed for Scenario 3

Rank	Idea	Cocreating count	No. of voters
1	There are valid reasons why people shouldn't base their education decisions on the results of a predictive model. If fees are based on the	1	7

	predictive model, they will be forced to change their decisions. For example, the model may be inaccurate.		
2	Reduces education to a purely economic argument, and also suggests a priority for vocational/field-orientated training when hiring trends suggest a growing demand for more diverse skillsets (such as the current shift from a pure STEM focus to a more diverse educational background).	1	7
3	People may have different preferences over the monetary and non-monetary rewards from education or people may have information about their individual circumstances that the model doesn't take into account.	1	7
4	The predictive model may produce "repugnant" results. For example, women may have a lower probability of finding a job in male dominated fields, not because women have lower ability, but because of the field's culture or discrimination. Basing course fees on the predictive model would reinforce these barriers.	4	8
5	Can statistical models predict for career paths that do not exist yet, or the multiple career paths an individual will likely have in their lifetime?	1	7
6	Training at a higher level can take 3-5 years. Managing demand in this way would assume that the economic situation could be predicted that far into the future. Varying the true cost would be a form of subsidy which itself would skew the market?	1	8
7*	Using the integrated data to analyse courses taken versus further employment and employment earnings is very interesting and is currently being analysed in the Student Loans project. This may be useful in advising students when they are considering taking a course but it should not influence how much the course should cost, or if the courses should exist at all. Not all courses are taken for monetary gain.	1	1

* An idea that was recently submitted and has yet to receive a sufficient number of votes to be properly ranked

Table 18: Constraints contributed for Scenario 3

Rank	Idea	No. of voters
1	make data available at both micro and macro level. Micro should be based on a model of individual choice. At micro level it allows students to make choices with more information about how their choices now may lead to jobs/vocations in future; It doesnt say students cant choose to go into their courses of choice - but choices arent always costless At macro level it allows decision makers to make more accurate determinations of future demand - meaning we actively recruit for future needs. The question implies that lower future demand could lead to higher course costs; the inverse also applies (future demand leads to lower course costs).	3
2	We struggle to match birth rates to the need for additional schools and teachers five years out. It is extremely difficult to predict what the skill sets we will need in the future and education tertiary education is not the only route to success	3

3	Make the model's predictions publicly available, but don't base course fees or admissions on the predictions. This lets individuals weigh the information provided with the model with other factors, e.g. private information about their abilities that the model doesn't have, personal preferences.	3
4	I am still concerned by the linear "X course = Y job" implicit in this. Some courses open a huge variety of career options - others are more aligned to a narrower professional identity (such as accountant). Also, education isn't only to produce workers. It should also produce citizens, creators, critics, and widen our collective knowledge and insight.	3
5*	Make the integrated dataset available to researchers who can build a variety of different models. I see what the model in the scenario does as very very hard. It is making predictions about the effect of a particular course on a particular student's job prospects. Such precise predictions are likely to be inaccurate. "One model to rule them all" is unlikely to work. Instead, researchers could derive less precise but more reasonable insights from the dataset. For example, it could be STEM students who didn't do Maths in secondary school are more likely to fail courses. In response, universities could put more resources into entry level Maths courses.	2

* An idea that was recently submitted and has yet to receive a sufficient number of votes to be properly ranked

Table 19: Benefits contributed for Scenario 3

Rank	Idea	No. of voters
1	Students would have a wider set of information available to them before making costly education decisions.	3
2*	We have had this ability for many years. It is not the data that is the issue, but the socio-political context that drives the decision making.	2
3*	managing courses and course numbers is sensible . Using the information to determine the costs of courses is not acceptable	2

* An idea that was recently submitted and has yet to receive a sufficient number of votes to be properly ranked