

In Confidence

Office of the Minister of Statistics

Chair, Cabinet Government Administration and Expenditure Review Committee

Data and Statistics Legislation: Paper 3 – Research and Analysis

Proposal

1. This is paper three in a suite of four papers proposing new data and statistics legislation. It seeks Cabinet agreement to policy proposals for sharing, integrating and accessing data for research and analysis.

Executive Summary

2. Integrated data (data brought together from a range of sources and linked using common variables) is a critical tool in the government's efforts to reduce poverty, improve health, and support a growing, inclusive economy. It enables stronger evidence-based insights and decision-making, and improves measurement of the effectiveness of government initiatives and services.
3. Because of the strong safeguards and protections in the Statistics Act 1975 (the Act), Stats NZ has been able to provide world-leading integrated data services to support public interest research and analysis to guide decision-making and generate new knowledge and insights. However, Stats NZ must rely on provisions designed for statistical production, to also guide decision-making about access to integrated data.
4. New data and statistics legislation will continue to provide necessary safeguards to protect people, communities and organisations represented in the data from unauthorised and inappropriate use. But, in addition, it will clearly specify the requirements that must be satisfied before access can be granted and include requirements for transparency about what data is being shared, by whom and for what purpose.

Background

5. Paper 1 summarises the previous Cabinet Committee and Cabinet consideration of the review of the Act and the 2018 public consultation on high level proposals in *Towards New Data and Statistics Legislation: Public discussion document*.

The value of integrated data for research and analysis

6. Data integration involves combining or linking data from different sources by using variables they have in common such as business or personal identifiers. Integrated data provides a much richer picture of New Zealanders, our society, economy, and environment. Its use significantly increases the value that researchers and analysts can obtain from data.

7. Integrated data is a critical tool in the government's efforts to reduce poverty, improve health, and support a growing, inclusive economy. It enables stronger evidence-based insights and decision-making, and improves measurement of the effectiveness of government initiatives and services.
8. Government agencies and external organisations share data they hold with Stats NZ so that it can be safely integrated and accessed for research and analysis in the public interest.¹ Approved researchers and analysts from government agencies may then access the integrated data to provide an evidence base for developing and evaluating policy. Research and analysis by government or other approved researchers and analysts leads to new insights and knowledge about things that are important to New Zealand, such as in the research examples below.

Agricultural productivity

Researchers at Motu Economic and Public Policy Research have used the Longitudinal Business Database (LBD) to explore [productivity of dairy and sheep/beef farms](#). Their research has shown that dairy farms achieve the greatest benefit from concentrating production in their primary activity. Sheep/beef farms benefit from adding other activities (eg forestry).

Well-being of Ngāi Tahu Whānui

Ngāi Tahu researchers are using data from the Integrated Data Infrastructure (IDI) and data held by Ngāi Tahu to understand how and why whānau may be at higher risk of serious health conditions. Other work is looking at how specific groups, such as [kaumātua](#), may be assisted to build economic resilience against poverty.

Natural environment and asthma

Researchers from Massey University have used the IDI to explore the connection between the [natural environment and asthma](#). They followed 50,000 New Zealand children born in 1998 through to 2016, and found that children who lived in greener areas were less likely to be asthmatic.

Making better study decisions

Career NZ's [Compare Study Options](#) tool helps young people make better-informed decisions about where their study choices can lead them. It compares earning and employment outcomes for different study options. The Ministry of Education created this tool by using combined student loan, tax, and education data in the IDI.

Economic costs of marital separation

A researcher at Auckland University of Technology has explored the economic consequences of marital separation, particularly for families with children. The [research](#) found that women are substantially worse off than men, and the negative impact persists for at least three years after separation. This research could be used to inform future child support policy in New Zealand.

Problem definition

9. Current legislative frameworks, including those in the Act and the Privacy Act 1993, support data sharing, integration and access for research and analysis, but they do not:
 - 9.1. expressly recognise and respect the Crown's responsibility to consider and provide for Māori interests in data and statistics;
 - 9.2. provide sufficient guidance for decision-making related to sharing, integrating and accessing data;

¹ See Appendix 1 for more information on Stats NZ's integrated databases.

- 9.3. require proactive transparency about who is using government-held data, how it is being used, what it is being used for, and how it is being managed.
10. The legislative policy proposals set out in this paper for sharing, integrating and accessing data address these issues. They reflect current best practice (domestic and international) including the policies and practices that underpin changes to legislation, for example, data protection legislation in other jurisdictions, the Digital Economy Act 2017 (UK) and the Australian National Data Commissioner's work to develop new data availability and transparency legislation.
11. They also reflect the framework developed by Stats NZ in partnership with Māori that applies a Te Ao Māori lens to managing access to integrated data (the tikanga framework).

Accessing data for research and analysis

12. A clear and robust legislative framework is required to effectively balance the benefits and risks associated with accessing government-held data for research and analysis.
13. The framework has been designed after consideration of current legislative provisions in the Statistics Act and the Privacy Act 1993, the tikanga framework, and international best practice for access to data held by national statistical offices or brought together at national data centres replicating the conditions of a national statistical office.
14. The legislative framework should incorporate safeguards that address each of the following risk factors:
- 14.1. research and analytic purposes – what the data can be used for;
 - 14.2. researchers and analysts – who can access the data;
 - 14.3. access environment – how the data can be accessed;
 - 14.4. data – what level of confidentiality is applied to the data before access;
 - 14.5. outputs – what results can be published.
15. Each of these factors refer to an independent but related aspect of risk. While each factor can be considered independently, all five need to be considered jointly to decide whether proposed data access is a safe arrangement. Safeguards can be increased or decreased as required to ensure that in total they provide for safe and appropriate data access.
- 15.1. If the data is lightly confidentialised (to reduce the risk of direct identification) other safeguards will be more stringently applied.
 - For example, data in Stats NZ's integrated databases has had identifying information removed (such as name and day of birth) and can only be accessed by approved researchers in a secure data lab environment.

15.2. If the data is strongly confidentialised (to protect even indirect identification of individuals or organisations), it can be accessed by a broader range of researchers and analysts with fewer additional safeguards applied.

- For example, data files that have been confidentialised using statistical methods such as aggregation, perturbation and suppression may be made available to approved researchers without requiring the data to be accessed in a secure data lab environment.

16. If the joint safeguards are not sufficient to manage the risks associated with the proposed data access, then access would not be granted. The safeguards that can be applied to manage each risk factor are discussed in the sections below.
17. Submitters on the 2018 public consultation support an access framework that recognises data sensitivity, likelihood of harm, and public expectations. Submitters agreed with the proposal that new legislation should provide more guidance on what should be considered when making decisions on access. Most submitters also supported increased transparency about access to government-held data, including who was accessing the data, how it was being used and for what purpose.

Research and analytic purposes – what the data can be used for

18. In the Act, the overarching test for what research and analytic purposes data can be used for is that it is in the public interest. However, the Act does not provide any direction on the types of matters that might usefully be considered before reaching a decision on public interest.
19. Research and analytic purposes that are in the public interest include purposes related to the functions and duties of government agencies such as developing and evaluating policy. It also includes a range of other research and analytic purposes where there is public good or benefit.
20. Retaining the public interest test was unequivocally supported by submitters during public consultation. Most submitters supported clarifying the public interest test considerations in new legislation including those that accounted for:
- 20.1. the nature and extent of any likely benefit (eg, contribution to improved wellbeing for Māori);
 - 20.2. the nature and extent of associated risks (eg, impact on public trust and confidence in the trusted use of data) or potential harms (eg, to individuals, communities, or organisations from research that takes a deficit view of the issue at hand or misrepresents a community);
 - 20.3. the context of data collection including the purpose of original data collection and information provided about possible future uses of data.
21. I propose that these considerations are adopted for the public interest test without limiting any other considerations from being taken into account. This will increase transparency in decision making and provide greater certainty both to researchers, analysts, and to those represented in the data.

22. I am not proposing that any group or sector be prevented from accessing government-held data as there can be significant public benefit from research and analysis that may also have private or commercial benefit. My view is that the requirement that the research and analysis be in the public interest, the application of the safeguards in paragraph 14, and the requirements for transparency will ensure any risks and harms are appropriately taken account of.
23. If data access for research and analytic purposes is contrary to an enactment, that limitation should prevail, and the public interest test will not be able to be satisfied. Otherwise, this safeguard is absolute. If the research or analytic purpose is assessed as not being in the public interest, access will not be granted.

Researchers and analysts – who can access the data

24. The framework should also continue to ensure that only ‘appropriately qualified’ researchers and analysts have access to the data. The appropriateness of the researcher or analyst will depend on whether they:
- 24.1. have the experience, knowledge and skills to access and use the data for the research or analytic purpose including the ability to work with data in ethically and culturally appropriate ways;
 - 24.2. can be trusted to use the data appropriately and protect the security of the data;
- and if the circumstances require, should include whether they can demonstrate:
- 24.3. engagement with the data provider(s);
 - 24.4. connections to, understanding of, or support from the communities the data relates to eg, iwi and Māori organisations, Pasifika or disability communities.

Access environment - how the data can be accessed

25. Access should be provided in an environment that is appropriate to safeguard the data, taking into account the likelihood and impact of unauthorised use or disclosure (deliberate or accidental). I do not consider it necessary for new legislation to include specific detail on what is or is not an appropriate access environment although an assessment of appropriateness would likely include:
- 25.1. the physical location where the data will be accessed;
 - 25.2. security and technical safeguards (eg, secure login, user authentication);
 - 25.3. support systems, policies, processes or protocols to ensure data is used in culturally-appropriate ways.
26. As is the case currently for the integrated data at Stats NZ, researchers and analysts will still be able to access data from an overseas location if it is assessed as appropriate. For example, Sydney-based researchers for the Ministry of Social Development are currently accessing integrated data for a specific project via a secure data lab, but the data itself is still at Stats NZ. Security arrangements for the

data lab in Sydney are the same as the data labs in New Zealand and the researchers are experienced and trusted users of the integrated data.

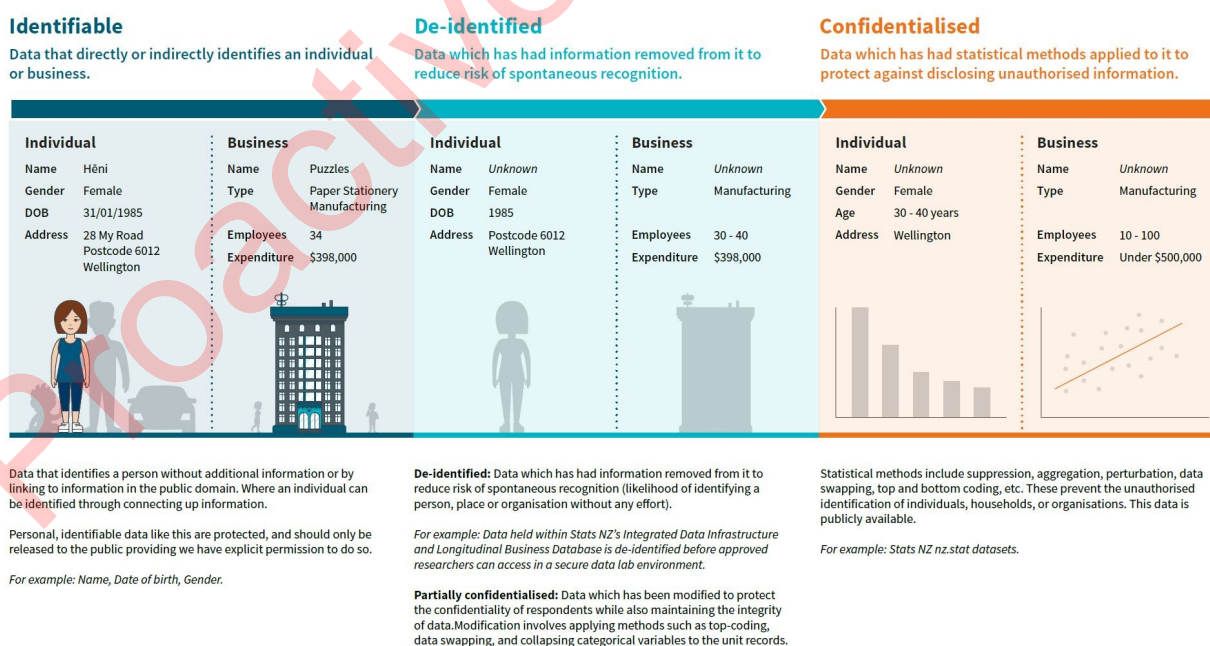
Data – what level of confidentiality is applied to the data before access

27. Appropriate and proportionate confidentialisation will need to be applied to data before it is accessed for research and analysis. In the case of personal information, at a minimum, direct identifiers (such as name and day of birth) will need to be removed before data is accessed for research and analysis.
28. Further treatment of the data may be required depending on:
 - 28.1. sensitivity of data including cultural sensitivity (eg, Māori knowledge and information relating to taonga, wāhi tapu and intellectual property);
 - 28.2. likelihood and impact of identification (including identification of individuals, collectives (eg, whānau, hapū and iwi), businesses, organisations and places (eg, location of endangered native species));
 - 28.3. other data that may be available to the researcher or analyst.
29. The following diagram shows the effect of different levels of confidentialisation on data relating to individuals and businesses.

Diagram 1: Degrees of identification in data

Degrees of identification in data

What do statisticians, data scientists and data analysts mean when they talk about confidentiality? How does identifiable data differ from de-identified or confidentialised information? Data identifiability is not binary. Data lies on a spectrum with multiple shades of identifiability. This is a primer on how to distinguish different categories of data in the NZ context.



30. In paper two, I set out the powers of the Government Statistician including developing and publishing guidance and standards. I expect these to include relevant guidance and standards that can be applied when de-identifying and confidentialising data.

Outputs – what results can be published

31. The current requirements in the Act and the Privacy Act relating to the publication of statistical or research outputs should continue to apply. These requirements prevent outputs from being published or disclosed in a form that could reasonably be expected to identify any individual, business or organisation unless an exception applies (eg, there is consent, the information is already publicly available).

Modern framework for sharing data across and with government

32. I have already discussed the value that can be created from research and analysis using shared and integrated data from government agencies and external organisations (eg iwi and Māori organisations, non-government organisations).
33. Submitters supported sharing data for research and analysis, with identified benefits including better informed policy development, less duplication, and a more complete picture of people and communities.
34. New legislation should clarify that government-held data can be used for research and analysis and shared across government and with government by external organisations for this purpose.²

Not all data is appropriate to share

35. I do not intend the new Act to override existing protections that limit data sharing for research and analysis in other enactments.
36. Where sharing or accessing data for research and analysis is contrary to an enactment, that enactment should prevail. For example, restricted records under the Births, Deaths, Marriages, and Relationships Registration Act 1995 are not shared with Stats NZ for inclusion in the Integrated Data Infrastructure (IDI).
37. Even when sharing is permitted, there may be limited benefit in doing so or the risks associated with sharing may be too high. Relevant considerations include:
- 37.1. the nature and extent of any likely benefit (eg, enabling research on specific populations or the relationship between different factors);
 - 37.2. the nature and extent of associated risks or potential harms;
 - 37.3. the integrity or quality of the data;
 - 37.4. the context of data collection including the purpose of original data collection and information provided about possible future uses of data.

² The Privacy Act permits the sharing and use of personal information for research and statistical purposes, but doesn't apply to economic or environmental data (unless also personal information).

38. New legislation should therefore require that agencies take into account the benefits and any risks associated with sharing the data before doing so.

When external organisations share data for research and analysis purposes

39. To avoid doubt the new Act should clearly state that the Government Statistician may agree conditions of access in relation to data shared by non-state sector organisations (eg, iwi and Māori organisations, non-government organisations or research organisations) for research and analysis alongside government-held data.
40. Specific access conditions may address concerns relating to the sensitivity of the data, commitments made to the people represented in the data, or other rights and interests that need to be respected (eg, research and analysis using mātauranga Māori data must benefit Māori because its value belongs to Māori).
- 40.1. For example, an iwi authority may agree to share their data with Stats NZ so that it can be linked with government-held data for research and analysis, provided that the iwi authority decides who can access their data and for what research and analytic purposes.
41. To address a further possible barrier to data sharing by non-state sector organisations, I propose that the Official Information Act 1982 does not apply to any data shared by those organisations.

Standards and guidance

42. I propose that the Government Statistician may develop and publish guidance and standards needed to support safe sharing and integration of data.

Accessing identifiable data when there is a legal authority to do so

43. The Act also prevents the Government Statistician from allowing access to the identifiable data that Stats NZ receives and integrates for use in research and analysis. In most cases, the appropriate course of action is for access to be via the data source agency.
44. However, in limited circumstances, it may be appropriate and more efficient to access the data directly from Stats NZ particularly where the relevant data is derived from multiple source agencies. The data available from Stats NZ may also be higher quality given its capacity and capability in accurately linking and/or processing the data (eg, data cleaning to detect and fix errors such as incorrectly coded variables).
- 44.1. For example, under the New Zealand Public Health and Disability Act 2000, mortality review committees (MRCs) have the power to access any information they require in order to perform their designated functions.
- 44.2. Most MRCs access identifiable or at least very detailed individual level data - the Family Violence Death MRC and Suicide MRC conduct in-depth analysis of the circumstances surrounding an individual's death across their life course, requiring considerable amounts of linked, identifiable data.

- 44.3. These committees have to apply separately to a number of agencies, and then integrate the data received themselves before being able to analyse it. This can take considerable time and effort, both for the committees and the agencies.
45. I propose that the Government Statistician have the power to authorise access to identifiable data held by Stats NZ for those organisations with legislative authority to access that same data at source or in another approved environment. Authorisation will only be made, following consultation with, and agreement from, the data source agency (or agencies).

Transparency

46. To ensure the transparency necessary for trusted use of government-held data, the new Act should require government agencies to publish summary information about:
- 46.1. what data is being shared, integrated and accessed for research and analysis (including any agreed access conditions)
 - 46.2. who is accessing the data (researchers and analysts)
 - 46.3. what the data is being used for (research and analytic purposes)
 - 46.4. how data is being accessed (access environment)
 - 46.5. data characteristics or limitations that may affect the quality of research and analysis.
47. Researchers and analysts accessing government-held data should also be required to publish or release:
- 47.1. results from their research and analysis; and
 - 47.2. research and analytic methodologies.
48. These requirements to publish information about data sharing, access, results and methodologies will apply unless there are reasons for not doing so (eg, information about research or analysis undertaken by a government agency during policy development may be withheld under the Official Information Act 1982).

Consultation


49. Officials have consulted with the Accident Compensation Corporation, Te Arawhiti, the Ministry of Business, Innovation and Employment, the Department of Conservation, the Department of Corrections, the Crown Law Office, the New Zealand Customs Service, the Ministry of Education, the Electoral Commission, the Ministry for the Environment, the Ministry of Foreign Affairs and Trade, the Government Communications Security Bureau, the Ministry of Health, the Ministry of Housing and Urban Development, the Inland Revenue Department, the Department of Internal Affairs, the Ministry of Justice, Land Information New Zealand, National Cyber Security Centre, the Office for Disability Issues, Oranga Tamariki, the Ministry for Pacific Peoples, the Ministry for Primary Industries, the

New Zealand Police, Te Puni Kōkiri, the Reserve Bank, the Ministry of Social Development, the Social Investment Agency, the State Services Commission, the Treasury and the Ministry for Women. Officials have also consulted with the Office of the Privacy Commissioner. The Department of the Prime Minister and Cabinet has been informed.

Financial Implications

50. The proposals set out in this package are fiscally neutral at a system/all of government level, and are expected to result in an unquantified net fiscal benefit to the Crown over time as the system becomes more reliable and efficient.
51. Costs associated with implementing updated legislative requirements at Stats NZ will be met within the agency's baseline.

Legislative Implications

52. Legislation is required to implement the proposals set out in this paper and the accompanying three papers, including replacing the Statistics Act with a new Data and Statistics Act.
53. 9(2)(f)(iv)

54. Should Cabinet agree to the proposals in this suite of papers, I anticipate being ready to introduce a new Bill to the House in the second half of 2020 with the aim of legislation passing in 2021.

Impact Analysis

55. A Quality Assurance Panel with representatives from the Ministry of Justice, the Regulatory Quality Team at the Treasury, and Stats NZ has reviewed the 'New data and statistics legislation' Regulatory Impact Assessment (RIA) produced by Stats NZ in February 2020. The Panel considers that the RIA meets the Quality Assurance criteria.
56. The panel notes the importance of the future work outlined in the RIA to confirm the level of penalties before the Bill is introduced.

Human Rights

57. The proposals in this paper are consistent with the New Zealand Bill of Rights Act 1990 and the Human Rights Act 1993.

Gender Implications

58. Improving the quality of, and access to, statistics and the data used for statistical purposes and for research and analysis, will contribute to monitoring the effectiveness of existing policies and other interventions, identifying emerging trends,

and providing information on the progress towards better outcomes for women, transgender, and intersex populations.

Disability Perspective

59. As better access to quality disability data is a priority in the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD), the New Zealand Disability Strategy 2016-2026, and Disability Action Plan 2019-2023, it is important for the disability sector that modernisation of data and statistics legislation creates the opportunity to progress access to quality disability data.

Publicity

60. I will make an announcement on the contents of the Data and Statistics Bill when the Bill is introduced. A commentary on the Bill will be released at this time.

Proactive Release

61. I propose that this Cabinet paper be released proactively on Stats NZ's website. The release will be subject to redactions as appropriate under the Official Information Act 1982.

Recommendations

62. The Minister recommends that the Committee:
1. **note** that government agencies and external organisations share their data with Stats NZ so that it can be integrated and safely accessed for research and analysis;
 2. **note** that research and analysis can lead to new insights and knowledge about things that are important to New Zealand;

Accessing data for research and analysis

3. **note** that the Statistics Act 1975 permits access to data for research and analysis provided certain conditions are met but does not expressly recognise the Crown's responsibility to consider and provide for Māori interests with respect to data and statistics, and lacks guidance for decision makers and requirements for transparency;
4. **agree** to a principle-based approach for accessing government-held data for research and analysis with safeguards that address each of the following risk factors:
 - 4.1. research and analytic purposes – what the data can be used for;
 - 4.2. researchers and analysts – who can access the data;
 - 4.3. access environment – how the data can be accessed;
 - 4.4. data – what level of confidentiality is applied to the data before access;

- 4.5. outputs – what results can be published;
5. **agree** that the research or analytic purpose must be assessed as being in the public interest, considering:
 - 5.1. the nature and extent of any likely benefit;
 - 5.2. the nature and extent of associated risks or potential harms;
 - 5.3. the context of data collection including the purpose of original data collection and information provided about possible future uses of data;
6. **note** that if the research or analytic purpose is assessed as not being in the public interest access will not be granted;
7. **agree** that researchers and analysts must be assessed as appropriate, considering whether they:
 - 7.1. have the experience, knowledge and skills to access and use the data for the research or analytic purpose including the ability to work with data in ethically and culturally appropriate ways;
 - 7.2. can be trusted to use the data appropriately and protect the security of the data;and, if the circumstances require, should include whether they can demonstrate:
 - 7.3. engagement with the data provider(s);
 - 7.4. connections to, understanding of, or support from the communities the data relates to;
8. **agree** that the environment in which the data is accessed must be appropriate taking into account the likelihood and impact of unauthorised use or disclosure (deliberate or accidental);
9. **agree** that appropriate and proportionate confidentialisation must be applied to data before it is accessed for research and analysis;
10. **note** that, for personal information, direct identifiers (such as name and day of birth) will need to be removed before data is accessed for research and analysis;
11. **agree** that researchers and analysts must take all reasonable steps to ensure that published or disclosed outputs no longer relate to identifiable individuals, businesses or organisations, or an individual, business or organisation that could reasonably be identified;
12. **agree** that the Government Statistician may develop and publish guidance and standards that can be applied when de-identifying and confidentialising data;

Sharing data across and with government

13. **agree** that identifiable data may be shared across and with government (by external organisations), and integrated for research and analysis, unless it is prohibited by the new legislation or contrary to another enactment;
14. **agree** that, when sharing is permitted by law, agencies must assess benefits and risks before sharing data, considering:
 - 14.1. the nature and extent of any likely benefit;
 - 14.2. the nature and extent of associated risks or potential harms
 - 14.3. the integrity or quality of the data;
 - 14.4. the context of data collection including the purpose of original data collection and information provided about possible future uses of data;
15. **agree** that the Government Statistician may agree conditions of access for research and analysis in relation to data shared by external organisations;
16. **agree** that data shared by external organisations will not be subject to the Official Information Act 1982;
17. **agree** that the Government Statistician may develop and publish guidance and standards needed to support safe sharing and integration of data;

Accessing identifiable data when there is a legal authority to do so

18. **note** that it may be more efficient for those with appropriate legal authority to access identifiable data from Stats NZ rather than from source agencies and such data may be higher quality given work to accurately link and/or process the data;
19. **agree** that the Government Statistician may authorise access to identifiable data held by Stats NZ by those with legislative authority to access that data at source or in another approved environment, following consultation with, and agreement from, the data source agency (or agencies);

Transparency

20. **agree** that government agencies be required to publish information about:
 - 20.1. what data is being shared, integrated and accessed for research and analysis (including any agreed conditions);
 - 20.2. who is accessing the data (researchers and analysts);
 - 20.3. what the data is being used for (research and analytic purposes);
 - 20.4. how data is being accessed (access environment);

- 20.5. data characteristics or limitations that may affect the quality of research and analysis;
- 21. **agree** that researchers and analysts accessing government-held data be required to publish or release (unless there are grounds to withhold):
 - 21.1. results from their research or analysis; and
 - 21.2. research and analytic methodologies;
- 22. **note** that published information about data sharing, integration, access, results, and methodologies may be in a summary form.

Authorised for lodgement

Hon James Shaw

Minister of Statistics

Date:

Stats NZ's integrated databases

Stats NZ is recognised as a world leader in the linking (or integration) of data for research and analysis, and protecting the identities of individual people and organisations. Stats NZ has two large integrated databases ([find out more about integrated data at Stats NZ](#)):

- Integrated Data Infrastructure (IDI) – contains data about people and households
- Longitudinal Business Database (LBD) – contains data about businesses.

Figure 1: Types of data in Stats NZ's IDI and LBD

