Stats NZ
Tatauranga Aotearoa

Integrated Data Infrastructure

Overarching privacy impact assessment

New Zealand Government

Citation
Stats NZ (2017). *Integrated Data Infrastructure: Overarching privacy impact assessment*. Retrieved from www.stats.govt.nz.

ISBN 978-1-98-852818-2 (online)

Published in July 2017 by
Stats NZ Tatauranga Aotearoa
Wellington, New Zealand

Contact
Stats NZ Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4600

www.stats.govt.nz

Contents

Purpose of this overarching privacy impact assessment

This privacy impact assessment (PIA) provides an overarching, systematic evaluation of privacy and other risks associated with integrating data into the Integrated Data Infrastructure (IDI), and how these risks are being managed. This PIA also summarises some of the expected benefits of the IDI. This document is relevant to all datasets in the IDI and should be used in conjunction with the IDI PIA extensions. The next section (Overview of privacy impact assessments for the IDI) explains the differences between this document and the PIA extensions.

See Integrated Data Infrastructure extension: Privacy impact assessment (7th ed) for information on datasets integrated before July 2016.

Stats NZ stewards the IDI, which is a centralised collection of administrative and survey data that is de-identified[1] and made available to approved researchers. The IDI provides a research database of longitudinal microdata about individuals, households, and organisations. Researchers can access the database to answer research, policy, and evaluation questions to support informed decision making that leads to positive outcomes for New Zealanders.

The PIA assists Stats NZ to:

- assess and minimise impacts on privacy
- assess if a proposed data integration is consistent with relevant legislation, including the Privacy Act 1993
- assess if a proposed data integration is consistent with Stats NZ's data integration guidelines.

Stats NZ is committed to conducting data integration activities transparently. We take steps to tell the public why and how we integrate data.

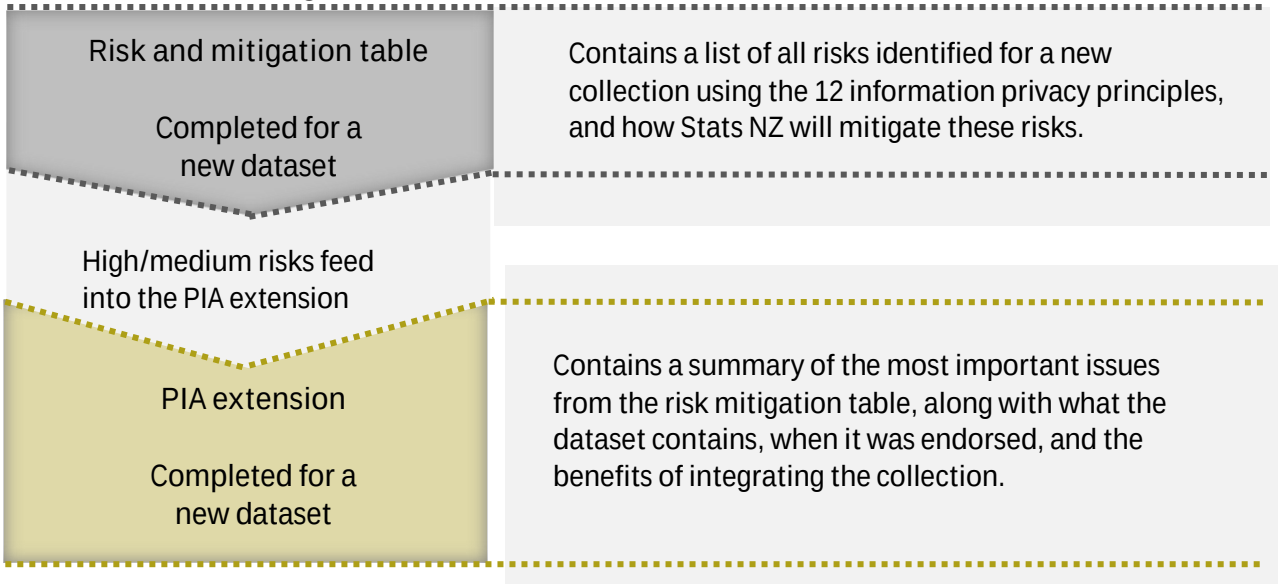
Overview of privacy impact assessments for the IDI

General privacy impact assessments for the IDI

Overarching PIA (this document) Applicable to all datasets in the IDI	Contains an overall evaluation of the privacy risks and benefits associated with integrating data into the IDI. When a new dataset is added to the IDI, a PIA extension and risk mitigation table are published separately. These documents should be read in conjunction with the overarching PIA.

[1] The process where information that could directly identify a person or organisation, such as names, addresses, or passport numbers, are removed or encrypted.

Process for creating a PIA extension for a new dataset



About the Integrated Data Infrastructure

Data integration is when data from two or more sources, which were originally collected for different purposes, are linked together.

The IDI is a special instance of data integration that has been approved, and is funded, by the New Zealand Government. It is made up of a series of datasets from different agencies that have been integrated. Individuals from single data sources are linked and then de-identified. The final dataset contains rich information about groups of people and their experiences over time. Researchers use integrated data to gain more insight into areas that will improve the social and economic outcomes of New Zealanders.

Researchers from across New Zealand can apply to access the IDI, subject to the strict requirements in the Statistics Act 1975 and in Stats NZ's policies and procedures. Researchers can only access the data they need to answer their research questions.

The infrastructure allows for security and privacy of individuals' information through de-identification of data and effective access and confidentiality controls.

The IDI has some key features that could affect its privacy risk without good stewardship. These are:

- The data integrated into the IDI will be held long-term.
- The IDI brings together information about different aspects of individuals' lives.
- Datasets are de-identified after inclusion in the IDI and before they are made available for statistics or research. However, Stats NZ will retain the identifying information separately, under strict conditions, so it can properly produce, maintain, and enhance the IDI, for example, in the context of adding further data to the IDI.

Four key privacy risks have been identified:

- individuals being re-identified in the data
- unfavourable public perception of the data integration

- inability to maintain data security
- data being used for non-approved purposes.

These risks have been classified using the Stats NZ rating tool. This rating tool is compliant with AS/NZS ISO 31000:2009 Risk Management – Principles and guidelines.

Risk	Likelihood of occurring	Consequence if it did occur	Mitigated risk rating
Individuals being re-identified in the data	Unlikely	Major	Medium
Unfavourable public perception of the data integration	Possible	Moderate	Medium
Inability to maintain data security	Unlikely	Severe	High
Data used for non-approved purposes	Rare	Major	Low

Summary of privacy risks and mitigations analyses these risks in more detail.

Expected benefits of the IDI

The IDI is a rich research database that can help decision makers find answers to complex issues. Specific benefits of the IDI include:

- meeting current demand for integrated research data that allows finer-grained statistical analyses and more detailed investigations. For example, investigating health status factors that fall outside the health system, such as economic (income, deprivation), environmental (housing, air quality, exposure to hazards) or lifestyle (smoking, drinking, exercise) factors
- allowing for the creation of additional statistics and research, to tell a richer story of people and organisations in NZ with no additional respondent burden
- enabling social policy experts to understand high needs and risks in new ways in order to better target interventions to achieve results
- providing opportunities for collaboration to address cross-sector issues, including the opportunity to better understand the interactions between different government services. For example, interactions between the health and welfare systems to develop better policies and practices for the prevention and management of different health conditions
- improving access to information sources by government and non-government researchers
- providing longitudinal data that enables researchers to tackle previously unanswerable questions
- encouraging wider use and re-use of data across government
- building capability in data integration expertise and in the analysis of integrated data, both within Stats NZ and across the Official Statistics System (OSS), by working in a unified environment.

Criteria for decision making

When weighing up the risks and benefits we refer to Stats NZ's data integration guidelines and the information privacy principles in the Privacy Act 1993 to aid in our decision making.

Principles for data integration

The principles for data integration define the key considerations that guide all Stats NZ's decisions on integrating data. It applies to integration of all personal data that we undertake for statistical or related research purposes, including for the IDI.

We will consider integrating data when all four of the following principles are met:

1. The public benefits of integration outweigh both privacy concerns about the use of data and risks to the integrity of the statistics system, the original source data and/or other government activities.
2. Integrated data will only be used for statistical or research purposes.
3. Data integration will be conducted in an open and transparent manner.
4. Data will not be integrated when an explicit commitment has been made to respondents that prevents such action.

These principles ensure that the proposed data integration is in keeping with relevant legislation, including the Privacy Act 1993, the Statistics Act 1975, the Public Records Act 2005, and any other legislation relevant to the source datasets.

The decision to integrate datasets, which in the case of the IDI is made by the Government Statistician, will be based on a careful examination of the issues, including undertaking a PIA when creating the business case for the addition of new data.

Information privacy principles

The Privacy Act 1993 contains 12 information privacy principles that set out how agencies may collect, store, use, and disclose personal information:

Principle 1: Purpose of collection of personal information.

Principle 2: Source of personal information.

Principle 3: Collection of information.

Principle 4: Manner of collection of personal information.

Principle 5: Storage and security of personal information.

Principle 6: Access to personal information.

Principle 7: Correction of personal information.

Principle 8: Accuracy of personal information to be checked before use.

Principle 9: Personal information not to be kept for longer than necessary.

Principle 10: Limits on use of personal information.

Principle 11: Limits on disclosure of personal information.

Principle 12: Unique identifiers.

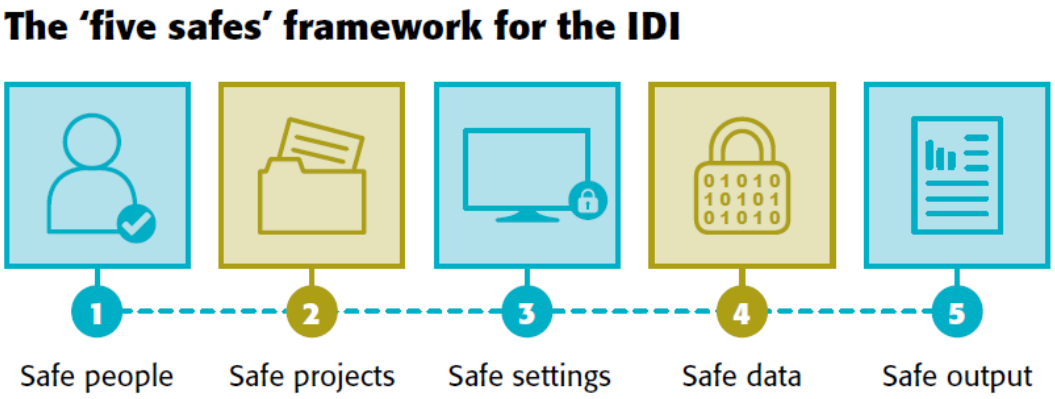
Read Information privacy principles on the website of the Office of the Privacy Commissioner for more detailed information about these 12 principles.

Management of the IDI

Stats NZ has a robust set of standard procedures that we use to manage and protect the IDI. See the outline of Stats NZ's security, access, and confidentiality systems.

'Five safes' framework for the IDI

The IDI uses a 'five safes' framework to ensure that access to data is provided based on the five conditions outlined below. This framework is based on international best practice.



1. Safe people – researchers can be trusted to use data appropriately and follow procedures. Before accessing data researchers must:

- pass referee checks
- sign a declaration of secrecy under the Statistics Act 1975
- sign a researcher undertaking
- attend confidentiality training
- follow Stats NZ rules and protocols.

Consequences for breaking protocols include being banned, blacklisted, or prosecuted.

2. Safe projects – the project has a statistical purpose and is in the public interest. Research is restricted to the analysis of groups, not individuals. This means that the research is focused on finding solutions to issues that are likely to have a wide public benefit.
3. Safe settings – security arrangements for data access are designed to prevent unauthorised access to the data.
   Data can only be accessed through a secure Data Lab environment. Computers are not connected to a work network and only Stats NZ staff can release data to researchers.
4. Safe data – the data inherently limits the risk of disclosure.
   We de-identify data, which means we remove personal identifying information such as names and addresses, and encrypt (ie replace with another number) identifiers such as IR and NHI numbers. Researchers get access only to the data relating to their research.
5. Safe output – the statistical results produced do not contain any information that could lead to the identification of an individual or entity.
   Researchers must confidentialise output before it can be released from the Data Lab, and

Stats NZ staff check results for confidentiality to ensure individuals cannot be identified. All output must go through a two-phased confidentiality checking process. See Microdata output guide for the methods and rules researchers must use for confidentialising output produced from Stats NZ's microdata.

Some of the important measures Stats NZ uses to support the 'five safes' framework are as follows:

Security measures

Stats NZ's standard security measures and protocols govern the management of the data in the IDI. We are required to comply with the confidentiality provisions of the Statistics Act 1975 and also Security in the Government Sector (SIGS) protocols.

Approved researchers from outside Stats NZ are able to access the IDI through a secure Data Lab[2] or, for a limited number of approved researchers, via remote access from their place of work. This remote access service allows researchers to access data via a secure data link to Stats NZ. No data can be downloaded, printed, or emailed. Regular security audits of these sites are conducted.

Stats NZ has well-established policies, procedures, and systems in place to ensure adequate measures of physical and electronic security. These include:

- Physical security systems control entry to premises and sections of premises to authorised persons.
- Security controls must be in place for remote Data Labs.
- Visitors to Stats NZ are subject to registration and supervision procedures and systems that ensure their activities are confined to legitimate business.
- Access to data is restricted on a 'need to know' basis.
- Access to our IT systems requires a valid user ID and password.

The Stats NZ security office actively audits and reviews security processes, and addresses new and emergent threats. Additional security arrangements for the IDI include:

- All identifying information used for processing the IDI is held on a secure, dedicated server, separate from data accessed for research and evaluation purposes.
- All data collections and associated electronic workspaces are secured (access is only authorised for those project personnel who need to access data for specific tasks and to selected IT administrators who are required to maintain the IT system).
- Data is only made available to approved researchers working on approved projects.
- Regular security audits check that only authorised individuals have access to the data.

Managing access to the IDI

Access to the IDI is strictly controlled:

- All research proposals are assessed using Stats NZ's microdata access guidelines.

[2] Currently located in each of Stats NZ's offices. Additional secure data labs have been established since 2014, with data access enabled via the remote access service.

- Data is only used for approved statistical or research purposes for the public good.
- Access is restricted to only those datasets required for each purpose.

Confidentiality measures

Confidentiality refers to the legal obligation Stats NZ has to protect information provided by individuals and organisations. Stats NZ has a strong culture of ensuring the confidentiality of information entrusted to us.

The IDI has a specific set of confidentiality rules that all researchers must comply with before any data is released. The confidentiality rules ensure individuals, households, and individual organisations cannot be identified. Any data released from the Stats NZ secure microdata access environment must undergo strict confidentiality checks before release.

The custodian of the integrated dataset is the manager of the team who oversees the day-to-day running of the IDI.

Datasets in the IDI

A prototype of the IDI was developed in 2012 and government invested in its expansion in 2013. New datasets are prioritised for inclusion by a cross-sectoral advisory group. Each dataset is then subject to a careful assessment of privacy and other risks, before being added to the IDI. The assessment of these privacy risks are included in Privacy analysis of data integration in this document and the PIA extensions. Stats NZ anticipates that up to eight datasets could be added yearly.

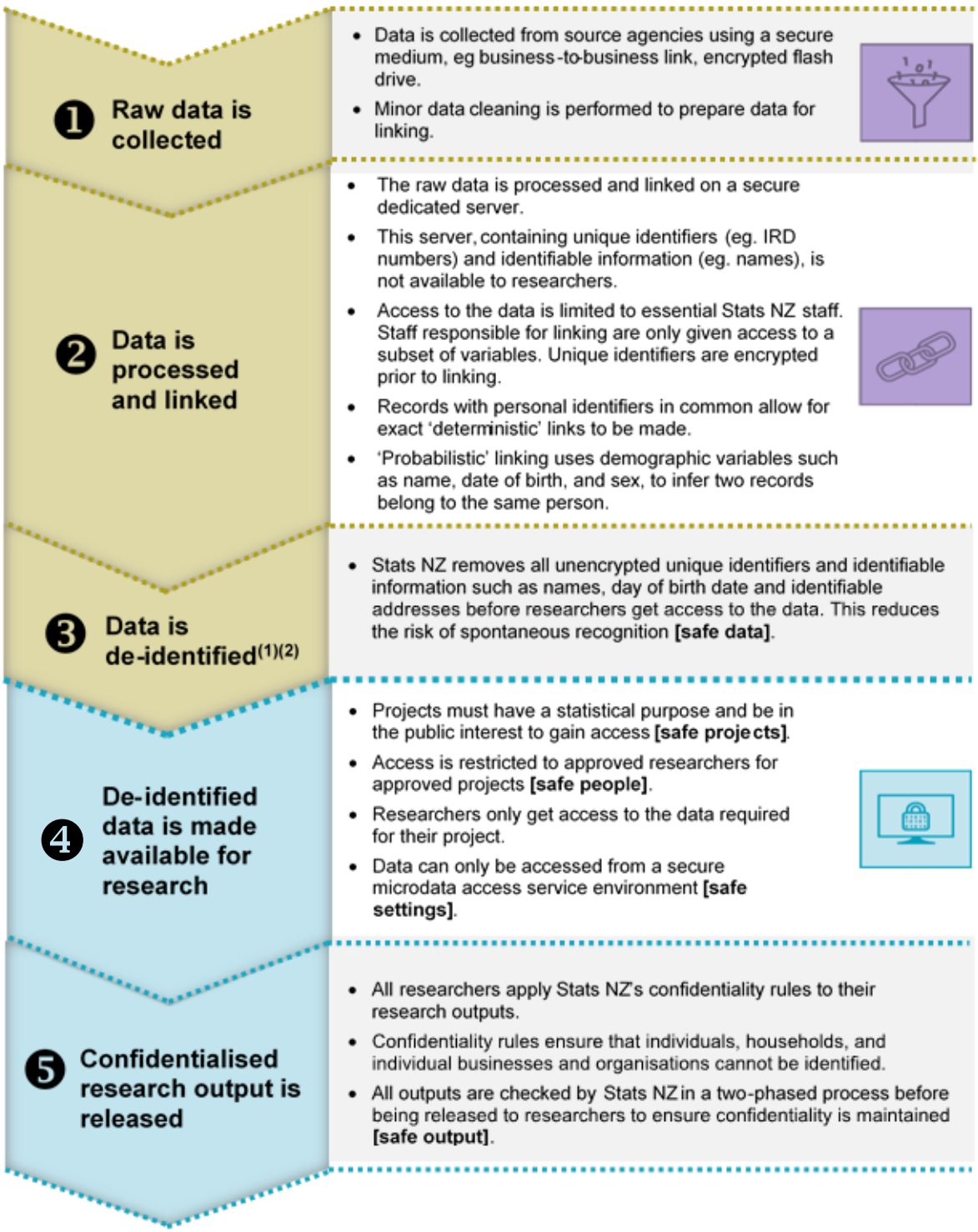
Overview of the datasets that are currently available in the IDI.

Variables in the IDI

Data dictionaries are available for the datasets in the IDI. Data dictionaries give information on the variables contained in the datasets – including technical information and descriptions. They contain information on coverage, methodology, and data quality.

Data integration information flows

This diagram shows the flow of information through the IDI system, and how personal information is collected, stored, accessed, and used at each stage.



1. A list of variables that are removed, encrypted or used for linking can be found in appendix 2.

2. De-identified: the process of removing information from microdata to reduce risk of spontaneous recognition. It will typically include removing names, exact dates of birth or death, and exact addresses. For more definitions see the glossary section.

Conclusion

The IDI integrates data from various agencies.

Four key privacy risks have been identified for this data integration. These are:

- individuals being re-identified in the data
- unfavourable public perception of the data integration
- inability to maintain data security
- data being used for non-approved purposes.

The procedures in place to mitigate the identified risks include the:

- application of the 'five safes' framework
- removal of unique identifying information from the datasets researchers can access
- strict confidentiality checking of the output of analysis
- stringent application process that researchers must follow to gain access to the IDI data, including the need to:
    - seek ethics approval for all projects wanting to access health data
    - meet strict access protocols in order to access the IDI data
    - demonstrate that the proposed research has a statistical purpose and is for the public good.

Due to the research value of the integrated dataset, an active dataset will be retained until the benefits no longer outweigh privacy concerns and risks.

If a review determines that the benefits no longer outweigh the risks, then the integrated datasets will be archived or destroyed (as required by the laws governing the archiving of government information).

Currently, the public benefits of the proposed data integration outweigh the potential privacy concerns. Well-established processes are in place to mitigate the risks identified, and these will continue to be reviewed and updated.

Stakeholders consulted

The following stakeholders have been consulted to prepare this privacy impact assessment:

Auckland City Missioner – Auckland City Mission
Legal counsel – Stats NZ
Manager, Labour and Income Statistics – Stats NZ
Manager, Population Statistics – Stats NZ
Privacy Officer – Accident Compensation Corporation
Privacy Officer – Department of Corrections
Privacy Officer – Department of Internal Affairs
Privacy Officer – Housing New Zealand
Privacy Officer – Inland Revenue
Privacy Officer – Ministry of Business, Innovation and Employment
Privacy Officer – Ministry of Health
Privacy Officer – Ministry of Justice
Privacy Officer – Ministry of Social Development
Privacy Officer – Ministry of Transport
Privacy Officer – New Zealand Police
Privacy Officer – New Zealand Transport Agency
Privacy Officer – Stats NZ
Respondent advocate – Stats NZ
Senior Policy Adviser – Office of the Privacy Commissioner

Legislation relevant to source data collections

Stats NZ must comply with legislation relevant to source data collections (such as the Tax Administration Act 1994 or the Health Information Privacy Code 1994).

Statistics Act 1975

Stats NZ operates under the Statistics Act 1975. As such, all data integration activities we undertake must adhere to the requirements of this Act, in particular the strict confidentiality and security provisions which operate to protect the data furnished to the department under the Act.

Legislation relevant to source data collections may result in the source agency seeking additional confidentiality measures, for example, Inland Revenue in terms of data provided under the Tax Administration Act 1994. However, no source agency legislation overrides the Statistics Act 1975 protections.

Sections of the Statistics Act that are of specific relevance to data integration activities include the following.

Section 2: Interpretation

'Official statistics' are statistics derived by government departments from statistical surveys and administrative data, the statistical analyses of which are published regularly, or could potentially be published regularly.

Section 3: Official statistics and coordination

Official statistics must be collected to provide the information that is required by the New Zealand Government, local authorities, and organisations to help make policy decisions. Official statistics must enable the government, government departments, local authorities, and the general public to understand economic, social, demographic, and other information of interest.

Section 4: Classes of official statistics

Stats NZ can require any person to provide information on any of the matters specified in the Act and other similar matters in the production of official statistics. This data integration includes the following matters:

- population and dwellings, migration (internal and external), vital and other demographic and social matters
- health, welfare, and morbidity
- cultural participation, education, and recreation
- law enforcement and the administration of justice
- matters relating to the social and physical environment
- labour and manpower, including conditions of employment; work descriptions; wages, including direct and indirect emoluments; hours of work and labour disputes
- accidents, including industrial injuries
- incomes, earnings, and related emoluments; expenditure and taxation

- household (including family) characteristics, conditions, and activities
- assets (including savings), liabilities, and wealth of persons, and undertakings
- economic, financial, production, and other matters relating to undertakings, including public administration, the Executive Government of New Zealand and local authorities; forestry, fishing, and trapping; agriculture; mines, quarries, and wells; manufacturing; construction; transportation, storage, and communications; electric power, gas, and water utilities; wholesale and retail trade; finance, insurance, and real estate; restaurants; hotels and accommodation; and other community, business, welfare, and personal services
- other similar matters, and such other matters as are prescribed by regulations under this Act.

Section 14: Duties of the Government Statistician

Subsection 14(c) requires that Stats NZ keeps the Minister of Statistics informed of the department's statistical projects, by explaining their purpose, scheme, methodology, and usefulness.

Subsection 14(d) requires that Stats NZ does not collect any information without the written permission of the minister.

Section 15: Independence of the Government Statistician

The Government Statistician has the sole responsibility for deciding the procedures and methods used to produce statistics. This includes data integration.

Section 37: Security of information

Data can only be accessed for statistical purposes and there are strict rules about how information is to be kept confidential.

Section 21: Declaration of secrecy

All personnel who access respondent information must keep this information secret for their lifetime.

Privacy Act 1993

The Privacy Act 1993 contains 12 privacy principles that set out how agencies may collect, store, use, and disclose personal information. Personal information relates to identifiable, living individuals.

All data integration activities we undertake must adhere to the requirements of the Privacy Act. The Act aims to uphold public trust in government by regulating the practice of data matching in the public sector.

Sections of the Privacy Act that are of specific relevance to data integration activities include the following.

Principle 1: Purpose of collection of personal information

Agencies must have a lawful purpose to collect information and collection must be necessary to fulfil this purpose. Stats NZ has a lawful purpose under the Statistics Act 1975 to integrate datasets to produce statistics.

Principle 2: Source of personal information

This principle requires information to be collected directly from the source. Stats NZ meets the exceptions that allow the organisation to use data collected by other agencies (eg from Inland Revenue). This is because the data will be used for statistical or research purposes [Principle 10 (f)(i)] and it will not be published in a way that could identify the individual [Principle 10 (f)(ii)].

Principle 5: Storage and security of personal information

This principle governs the way personal information is stored. It is designed to protect personal information from unauthorised use or disclosure. Stats NZ's standard security measures and protocols govern the management of the data in the IDI. We are required to comply with the confidentiality provisions of the Statistics Act 1975 and also Security in the Government Sector (SIGS) protocols. Stats NZ has well-established policies, procedures, and systems in place to ensure adequate measures of physical and electronic security.

Principle 12: Unique identifiers

This principle requires that an agency does not assign a unique identifier that has been assigned to an individual by another agency (eg an IRD number assigned by Inland Revenue). Current practice at Stats NZ is that unique identifiers assigned by other agencies are used at the time of construction of the integrated datasets, but are not retained permanently in the datasets. The original unique identifiers for persons represented in the original datasets are removed. The final datasets contain (de-identified) unit records identified only through a Stats NZ unique reference.

Public Records Act 2005

The Public Records Act 2005 sets out requirements for creating and maintaining adequate records of the business of public offices, including Stats NZ.

All source data for integration and resulting integrated datasets must be retained or disposed of according to the requirements of the agreement between Stats NZ and the Chief Archivist, authorised under section 20 of the Public Records Act. The following disposal schedules govern what data must be preserved, and what can be destroyed when no longer needed for statistical purposes:

- Statistical data, documentation, and metadata disposal schedule (DA379)
- 2006 Census of Population and Dwellings disposal schedule (DA439)
- 2011 Census of Population and Dwellings disposal schedule (DA440)
- Statistical schedules disposal schedule (DA271)
- Administrative Data disposal schedule (DA272)
- 2013 Census of Population and Dwellings disposal schedule (DA563)

Health Information Privacy Code 1994

The Health Information Privacy Code 1994 regulates the collection of health information by health agencies. The Code covers the collection of information, rights of individuals to access and correct information, retention of information, and limits on disclosure and use. The Code specifies the conditions in which disclosure and use are permitted, including for statistical or research purposes.

Sections of the Code that are of specific relevance to data integration activities include the following.

Rule 11: Limits on disclosure of health information

1) A health agency that holds health information must not disclose the information unless the agency believes, on reasonable grounds:

(b) that the disclosure is authorised by:

(i) the individual concerned; or

(ii) the individual's representative where the individual is dead or is unable to give his or her authority under this rule;

(c) that the disclosure of the information is one of the purposes in connection with which the information was obtained;

2) Compliance with paragraph (1)(b) is not necessary if the health agency believes on reasonable grounds that it is either not desirable or not practicable to obtain authorisation from the individual concerned and:

(c) that the information:

(iii) is to be used for research purposes (for which approval by an ethics committee, if required, has been given) and will not be published in a form that could reasonably be expected to identify the individual concerned.

The Ministry of Health is comfortable with Stats NZ's approach to ensure that information is correctly stored, used, and identifiable information deleted as soon as possible.

Births, Deaths, Marriages, and Relationships Registration Act 1995

The Births, Deaths, Marriages, and Relationships Registration Act 1995 details how life events are notified, registered, amended, and accessed. Sections that are of specific relevance to the provision of identity information for inclusion in the IDI include:

Part 9: Searches and disclosure of information

This part details who is authorised to access source documents and includes restrictions on searches, for example:

- Section 75A Subject of birth information, marriage information, civil union information, or name change information may search access register and request non-disclosure direction.
- Section 76 Restrictions on searches relating to adopted persons.
- Section 77 Restrictions on searches where error relating to person's sex has been corrected or sexual assignment or reassignment has been registered.
- Section 78 Restrictions on searches relating to new names of certain witnesses, etc.

These restricted records are not provided to Stats NZ for inclusion in the IDI.

Part 11A, Section 87A Provision of information to Stats NZ

This Section of the Act refers to what information can be made available to Stats NZ. It states that 'This Act does not limit the provision of information under the Statistics Act 1975'.

District Court Act 2016 and Senior Courts Act 2016

The Ministry of Justice can only share court information under specific authority.

Court information is shared with Stats NZ under two authorities: an Approved Information Sharing Agreement that authorises the sharing of 'permitted information' under section 237 of the District Court Act 2016 and section 174 of the Senior Courts Act 2016; and a specific authority from the Principal Judge of the Youth Court for sharing Youth Court Information.

Other legislation

Source data collection agencies may collect the data provided to Stats NZ for data integration activities under their own legislative framework. Their legislation also details any restrictions to data sharing, and the onus is on each source agency to ensure that in giving us data both the source agency and Stats NZ are complying with the relevant pieces of legislation. No other legislation can override the privacy and confidentiality protections of the Statistics Act 1975, but may require additional protections. If so, these additional protections will be detailed in a memorandum of understanding between Stats NZ and the agency.

An example of where extra protections are needed in order for Stats NZ to comply with source agency legislation is detailed in the Tax Administration Act 1994.

Section 81 of the Act imposes secrecy obligations on all employees of Inland Revenue. The secrecy obligations mean that Inland Revenue cannot disclose information to a third party, unless it is for the purpose of carrying into effect any of the Inland Revenue Acts or an exception (in section 81(4)) applies.

Section 81(4)(d) of the Act authorises Inland Revenue officers to provide information to Stats NZ staff. It also enables Stats NZ to produce statistics from tax information. All information supplied to Stats NZ is defined as restricted information, to which further secrecy requirements apply. These are outlined in section 87 of the Act and state that any person with access to restricted information shall certify that he or she has been shown, has read, and has understood the

provisions of section 87, and that, like Inland Revenue officers, they will also be subsequently bound to maintain and aid in maintaining the secrecy of all restricted information.

Glossary: Acronyms, initialisms, and definitions

Acronyms and initialisms

IDI – Integrated Data Infrastructure

IRD – Inland Revenue Department

OSS – Official Statistics System

PIA – privacy impact assessment

SIGS – Security in the Government Sector

Definitions

Administrative data: data collected by an organisation through its operations or provision of services.

Confidentiality: The principle that when people and organisations provide information to us, we should not disclose it to people who are not authorised to have access to it. Authorisation should ideally be by the person providing the information, but may also be through legislation.

Confidentialisation: The statistical methods used to protect against confidential information being disclosed to people who are not authorised to have access to it, in a way that could identify an individual, household, or organisation.

Data integration: Combining data from two or more sources.

De-identification: The process of removing information from microdata to reduce risk of spontaneous recognition. It will typically include removing names, exact dates of birth or death, and exact addresses. Also referred to as anonymisation.

Deterministic linking: Linking records belonging to the same unit through a unique identifier.

Longitudinal: Observing or examining a group of people or things over time.

Microdata: Data about individual people, households, or organisations. It may also be data about other characteristics of New Zealand such as geographical information.

Personal information: Information about a person that we should not disclose to people who are not authorised to have access to it. It is a subset of confidential information.

Privacy: The individual's rights relating to control of the provision, use, and disclosure of information about themselves, commonly called their personal information.

Probabilistic linking: Methodology based on the relative likelihood that two records belong to the same unit, given a set of similarities/differences between the values of the linking variables (eg name, date of birth, sex) on the two records.

References

Office of the Privacy Commissioner (1994). Health Information Privacy Code 1994 (updated October 2015). Available from www.privacy.org.nz.

Stats NZ (2012). Privacy impact assessment for the Integrated Data Infrastructure. February 2012. Available from www.stats.govt.nz.

Stats NZ (2016). Integrated Data Infrastructure extension: Privacy impact assessment (7th ed). July 2016. Available from www.stats.govt.nz.

Appendix 1: Privacy analysis of data integration

This section examines privacy risks associated with data integration in general. The four key privacy risks that have been identified for data integration are:

- individuals being re-identified in the data
- unfavourable public perception of data integration
- inability to maintain data security
- data used for non-approved purposes.

These risks are system wide and not specific to any dataset. From January 2017, detailed analysis of the risks associated with specific datasets are available in the PIA extensions.

Read the privacy impact assessment extensions.

These risks will be reviewed if there is a significant change to the current data integration process in the IDI.

Summary of privacy risks and mitigations

Individuals being re-identified in the data

Description of risk

The use of personal information could lead to individuals being identified. Due to the broad range of data in the IDI it is possible that, even though key identifiers such as names are removed, a researcher may spontaneously recognise an individual in the data due to the fact that they know a number of other facts about the person. It is also possible that human error may result in output being released that could identify an individual, household, or organisation.

The identification of individual information could lead to a breach in privacy. This could result in distress for the individuals involved and in loss of public confidence in agencies or service providers responsible for safeguarding personal information. A breach in privacy may impact the quality of information respondents provide to service providers and agencies across the OSS. In the health system, patients may feel that they cannot fully disclose information to their treating clinicians, which could result in serious harm to the patients.

A breach in privacy may also bring into question the process of linking data from unrelated sources and would affect the public's perception of Stats NZ as a trusted data collector/holder. Additionally, our OSS partners may not be willing to provide us with further administrative data if a breach occurred.

How the risk will be mitigated

Access to the incoming datasets with identifying information on them will be limited to the team overseeing the IDI linking. In compliance with the Statistics Act 1975, all identifying information will be removed from statistical datasets to prevent identification of individuals.

Researchers using the IDI data will only have access to the specific datasets required for their research questions. Processes are in place to assess a potential researcher's integrity and experience. Once researchers are approved they must complete training in applying

confidentiality methods and sign both a Declaration of Secrecy and a Researcher Undertaking before they can access the data. If researchers have been approved to access Inland Revenue data, a Tax Secrecy Declaration must be signed.

The Researcher Undertaking includes an agreement to:

- not attempt to identify particular persons or organisations
- not attempt to match the information with any other unit record level data source or list of persons or organisations
- not disclose, either directly or indirectly, information protected by the Statistics Act 1975 with any individual not approved by Stats NZ
- apply confidentiality measures to all output to ensure that no individual person or organisation can be identified
- take all reasonable steps to ensure that no unauthorised person can view the data.

In addition, a contract is signed between Stats NZ and the research agency undertaking the research. Among other things, this contract commits the agency to ensure their researchers follow all instructions and directions designed to protect confidentiality of data.

Stats NZ develops confidentiality measures for each data type. The techniques selected are data specific and may suppress information at different levels, in order to prevent identification of individuals, families, or organisations. Additionally, we will maintain confidentiality standards by checking research output before release, ensuring identifying information is not released.

Secure computer systems ensure that researchers accessing the data cannot print, email, or access the internet (this restricts downloading capability, access to file-sharing websites, and web-based email). Audit functions are used to monitor researchers' use of the systems. Research output can only be released by Stats NZ staff once confidentiality checked and approved for release.

Mitigated risk rating

Medium.

Likelihood of risk occurring

Unlikely.

Impact of risk occurring

Major.

Unfavourable public perception of the data integration

Description of risk

There may be unfavourable public perception towards the data integration and the public may not see this as an appropriate use of data.

Unfavourable public perception about the integration of data for the IDI may bring into question the legitimacy of linking data from unrelated data sources. This may impact on the IDI itself, and also on the reputation of Stats NZ and source agencies. This would affect the public's perception of Stats NZ as a trusted data collector/holder, and may affect the quality of information we

receive. For example, an individual may refuse to complete the census form or there could be a public campaign during the census field operation.

Unfavourable public perceptions may also result in a loss of public confidence in agencies or service providers responsible for the safeguarding of personal information. This may impact the quality of information respondents provide to agencies and service providers. Additionally, our OSS partners may not be willing to provide us with further administrative data.

How the risk will be mitigated

This risk will be minimised by:

- adhering to Cabinet directive [CAB (97) M31/14]
- consulting with the Office of the Privacy Commissioner
- being transparent about project objectives and processes
- consulting with key stakeholders
- complying with the Statistics Act 1975, the Privacy Act 1993, and other relevant legislation
- only granting access to data on a 'need to know basis'
- auditing access
- conducting research on public attitudes to data integration
- developing a communications plan that addresses any areas of concern and provides clear and accurate messages to the public
- operating within the parameters of what people consider to be acceptable use of the data.

Mitigated risk rating

Medium.

Likelihood of risk occurring

Possible.

Impact of risk occurring

Moderate.

Inability to maintain data security

Description of risk

People accessing the data that do not have permission to do so.

Unapproved access to the data would affect the public's perception of Stats NZ as a trusted data collector/holder, and may affect the quality of respondent information we receive. Additionally, our OSS partners may not be willing to provide us with further administrative data.

How the risk will be mitigated

Data transfer from the source agency to Stats NZ is strictly controlled, which minimises the risk as data is transferred. Additionally, the data will reside on a secure server within Stats NZ, which means individuals outside the team that oversee the IDI won't be able to access the data. Stats NZ

removes all unencrypted unique identifiers and identifiable information such as names, day of birth date, and identifiable addresses before researchers get access to the data. Researchers are required to complete a stringent application process before being granted access to the data. Approved researchers working on approved projects can access the IDI in a secure Data Lab, or via the remote microdata access facility, under strict controls. We are able to audit access to the IDI during all stages of the process.

Mitigated risk rating

High.

Likelihood of risk occurring

Unlikely.

Impact of risk occurring

Severe.

Data used for non-approved purposes

Description of risk

The linked data in the IDI is not being used for bona fide research or statistical purposes in relation to a matter of public interest, for example, it is being used to identify individuals or as the basis for case-management decisions about individuals.

Linked data in the IDI being used for non-approved purposes may not only impact on the IDI itself, but also on the reputation of Stats NZ and the source agencies. Any use of the data other than for approved research purposes would affect the public's perception of Stats NZ as a trusted data collector/holder and may affect the quality of respondent information provided to agencies and service providers. Additionally, our OSS partners may not be willing to provide us with further administrative data.

How the risk will be mitigated

Researchers adhere to a strict application process before being given access to the IDI data. The conditions of access include:

- being granted access for specific genuine research purposes only, for approved projects and approved researchers
- only being able to access specific de-identified datasets relevant to their research question
- being subject to formal approval processes by Stats NZ where each initial research request will be assessed by the Government Statistician
- a two-phased confidentiality process for all outputs leaving the secure data environment.

The dataset will be archived when risks outweigh benefits.

In addition to these checks, the confidentiality checks that are performed before release of findings ensure that the linked data in the IDI is used only for statistical research purposes.

Mitigated risk rating

Low.

Likelihood of risk occurring

Rare.

Impact of risk occurring

Major.

Appendix 2: Variables that are removed or encrypted and used for linking

Variables that are removed or encrypted

- first names
- middle names
- last names
- titles
- organisation names
- day of date of birth
- day of date of death
- day of date of disposal
- day of date last seen
- address information
- identity numbers such as Inland Revenue number or passport number.

Variables that are used for linking

In order to link individuals across different data sources, a number of variables are used to verify that a person is the same individual. Some variables are personal identifiers (such as name and date of birth) and some variables are given to the individual by the agency that they are interacting with (such as IRD number or passport number).

The list of variables used in the linking process includes:

- name(s) and any aliases
- date of birth
- sex
- ethnicity
- address
- source specific identifiers such as IRD number or passport number.

Appendix 3: Risk rating tool

Table 1
Risk consequence table

Rating	Consequence
Severe	Would require extensive senior management attention and diversion of resources to recover from the risk event.
Major	Significant senior management attention would be required to recover from the risk event.
Moderate	Management effort would be required to prevent the situation from intensifying. Changes to operating procedures would be required.
Minor	Management oversight would be required to ensure effectiveness and efficiency is maintained. Changes to operating procedures may need to be considered.
Insignificant	Management oversight might be required to ensure day-to-day, routine operations are not disrupted.

Table 2
Risk likelihood table

Rating	Likelihood of occurrence
Almost certain	The risk consequence will occur in most circumstances. 80–100% expectation in the next 12 months.
Likely	The risk consequence will probably occur. 50–80% expectation in the next 12 months.
Possible	The risk consequence is liable to occur. 30–50 % expectation in the next 12 months.
Unlikely	The risk consequence may occur at some time. 5–30% expectation in the next 12 months.
Rare	The risk consequence will only be realised in exceptional circumstances.

Table 3
Risk rating matrix

	Insignificant	Minor	Moderate	Major	Severe
Almost certain	Medium	High	High	Very high	Very high
Likely	Low	Medium	High	High	Very high
Possible	Low	Low	Medium	High	High
Unlikely	Very low	Low	Low	Medium	High
Rare	Very low	Very low	Low	Low	Medium

Compliant with AS/NZS ISO 31000:2009 Risk Management – Principles and guidelines