Statistics New Zealand
TATAURANGA AOTEAROA

Privacy impact assessment for the Integrated Data Infrastructure

New Zealand Government

Citation
Statistics New Zealand (2012). *Privacy impact assessment for the Integrated Data Infrastructure.* Available from www.stats.govt.nz.

ISBN 978-478-37749-1(online)

Published in February 2012 by
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

Contact
Statistics New Zealand Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4610
www.stats.govt.nz

Contents

List of figures

Figures by chapter

1 Introduction and overview

1.1 Introduction

This privacy impact assessment has been prepared for the Integrated Data Infrastructure (IDI). The IDI has the following components:

- iLEED (integrated Longitudinal Employment and Education Data)[1]
- migration and movements data
- Longitudinal Business Database (LBD) data.[2]

The assessment identifies privacy risks associated with this data integration and outlines the processes for managing these risks.

It was prepared in accordance with Statistics NZ Data Integration Policy (see appendix G) and the Linked Employer-Employee Data (LEED) Project requirements. The LEED Privacy Impact Assessment (see appendix A) states:

> Any proposal to link additional personal information to this framework would require separate approval including a further PIA and approvals by the Government Statistician and the appropriate data stewards, and may be subject to comment or other process by the Privacy Commissioner operating under the Privacy Act 1993.

The aim is to create a longitudinal integrated data infrastructure to allow linking of individual, household, and business-level data. This will enable statistical outputs on the transitions and outcomes of people through the secondary and tertiary education systems, the labour market, the benefit system, movements in and out of New Zealand, and links to business data. The infrastructure will further allow increased security, and flexibility to respond to changes and development in source-agency administrative datasets and Statistics NZ processes and outputs.

1.2 Overview

This report begins with a description of the legal context. It then describes the major elements of the infrastructure and the flow of information through it. Privacy issues associated with the integration, storage, and use of the source data are discussed, along with a description of the release and access practices to be used. This discussion includes an assessment of the risks inherent in these processes, and the privacy enhancement and risk management procedures that will mitigate them.

The report concludes with a brief discussion of the compliance mechanisms to be used to manage the integrated data infrastructure in the future.

---

[1]The iLEED component of the IDI consolidates data from the following existing integrated datasets:

- Linked Employer-Employee Data (LEED)
- LEED-Ministry of Social Development (MSD) benefit data
- LEED-Statistics NZ's Household Labour Force Survey (HLFS) data
- Student Loans and Allowances integrated dataset
- Employment Outcomes of Tertiary Education data.

[2] Any reference to the IDI or infrastructure throughout this document implicitly includes the iLEED, migration and movements, and LBD components.

1.3 Legal context

In the past, submissions to the government have pointed to privacy issues around agencies' data integration initiatives. Public confidence and acceptance are concerns that need to be properly managed. To address these concerns and the risks inherent in integration work, Cabinet agreed that:

> where datasets are integrated across agencies from information collected for unrelated purposes, Statistics NZ should be the custodian of these datasets in order to ensure public confidence in the protection of individual records. (CAB (97) M 31/14)

Statistics NZ has considerable experience complying with and applying the requirements of the Privacy Act 1993. Well-established practices ensure security and confidentiality of information. There is a strong culture within Statistics NZ around the confidentiality and privacy of information entrusted to us by individuals and firms.

Statistics NZ's existing microdata access protocols (see appendix H) are designed to protect the confidentiality of individual information, which will guard against confidentiality breaches. The data integration policy (appendix G) exists to address privacy concerns raised by the process of integration.

Security, privacy, and confidentiality rules are operationalised through Statistics NZ policies, including:

Privacy:

- Data Integration Policy

Security:

- Security framework

Confidentiality:

- Confidentiality Standard for Microdata access
- Confidentiality Standard for Census
- Confidentiality Standard for Social Collections
- Confidentiality Standard for Business Collections
- Confidentiality Principles of the Statistics Act 1975
- Policy for the Release of Statistics
- Policy for Publishing Official Statistics
- Microdata Access Protocols
- Confidentiality rules for release of aggregated census data.

The Statistics Act 1975 and Privacy Act 1993 form a framework to protect information about individuals when used for statistical or research purposes. The Privacy Act 1993 protects individuals (living natural persons), while the Statistics Act 1975 has a wider security coverage including natural or deceased people.

The Statistics Act 1975 (the Act) regulates the collection of information, whether from statistical surveys or administrative records, for use in producing official statistics by government departments. This Act includes strict provisions to protect the security of collected information and to prevent the release of identifiable information about an individual or business (information is still considered identifiable when identifiers such as names are removed, but where a third party could identify that the information relates to a particular person). Section 37 of the Act states that information provided under the Act

shall only be used for statistical purposes, the nature of which will include research undertaken by Statistics NZ employees, secondees, and bona-fide research undertaken in the Statistics NZ Data Lab.

In accordance with the Statistics Act 1975, Statistics NZ ensures that information it receives is kept securely, access is restricted, and any publication avoids disclosing identifiable information about a person, household, or business. All data held by Statistics NZ is subject to these strict provisions.

The Privacy Act 1993 protects information about an individual and applies to every agency that deals with personal information. Twelve information privacy principles in the Privacy Act 1993 provide a foundation, which governs the protection of privacy for the collection, use, disclosure, storage, and access to personal information. It should be noted that Statistics NZ has special provision under the Privacy Act 1993 to use personal information to produce statistics and that data will not be released in a way that could identify the individual [10 (f) (i) and (ii)].

Under the terms of the Privacy Act 1993, the Privacy Commissioner can conduct an independent review in the event of a complaint. Because of this, the commissioner is not able to approve proposals like these in advance. However, the commissioner is able to signal any practices that are not permitted under the Privacy Act 1993 or that might pose a problem of perceived privacy risks. Statistics NZ has taken the position that any such concerns, even of perception, should be addressed in an appropriate and defensible manner before integration begins. The intention is that Statistics NZ should take all necessary steps to comply with both the spirit and letter of the Privacy Act 1993.

Compliance with the Statistics Act 1975 and the Privacy Act 1993 ensures that the privacy concerns of individuals are not set aside for the potential value and perceived benefits of record linking. Business data used in the infrastructure is covered under other existing Statistics NZ rules.

The infrastructure will be carried out in line with legal requirements under the Statistics Act 1975, the Privacy Act 1993, and relevant sections of the Tax Administration Act 1994 and the Social Security Act 1964.

All data integration requires some form of identifying information to ensure the data is effectively linked. While a variety of identifying information can be used for linking, the use of unique identifiers generally gives the most accurate outcome. This means that particular care is needed to comply with information privacy principle 12, which prevents an agency from assigning a unique identifier used by another agency. Statistics NZ has worked in conjunction with the Office of the Privacy Commissioner (OPC) to ensure that its data integration work complies with this principle.

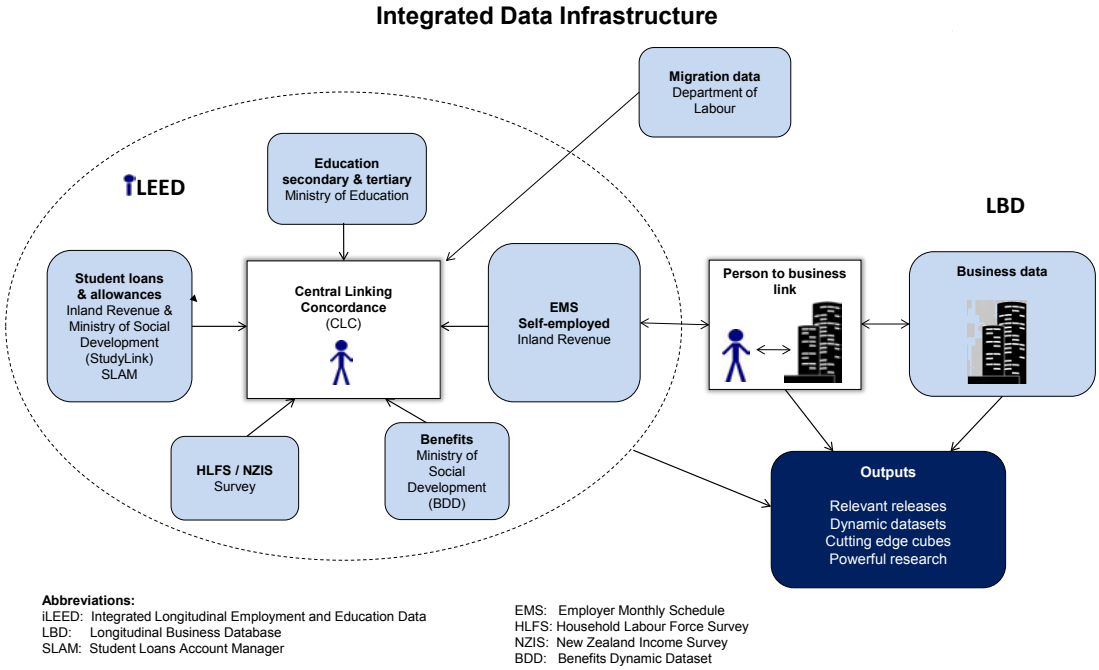
2 Description of infrastructure and information flows

2.1 Infrastructure description

The IDI will allow linking of individual and business-level data. This will enable statistical outputs on the transitions and outcomes of people through the secondary and tertiary education systems, the labour market, the benefit system, movements in and out of New Zealand, and links to business data.

The infrastructure will also support the security of an individual's information, through effective access controls, and flexibility to respond to changes and developments in source-agency administrative datasets and Statistics NZ processes and outputs.

Figure 1

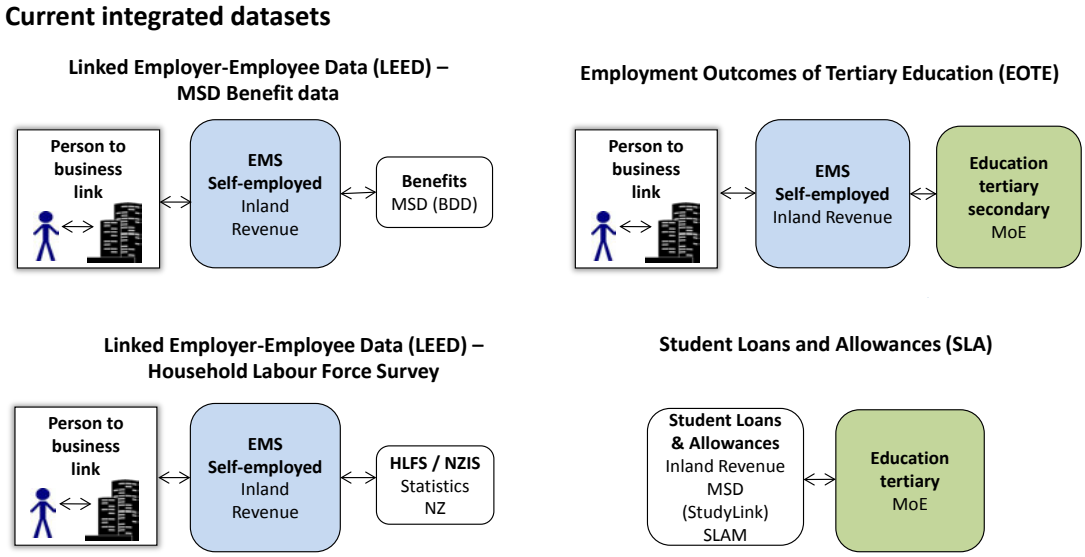


Statistics NZ currently maintains several different integrated datasets. There is considerable duplication across these datasets.

- Linked Employer-Employee Data (LEED) integration with Ministry of Social Development (MSD) benefit data. LEED combines Inland Revenue tax data with Statistics NZ Business Frame information and links through to Statistics NZ business information in the prototype LBD.
- The Student Loans and Allowances (SLA) integrated dataset, which combines tertiary education data with MSD and Inland Revenue student loans and allowances information.
- Employment Outcomes of Tertiary Education (EOTE) data, which combines tertiary education data with LEED and links through to Statistics NZ business information in the prototype LBD.
- LEED integrated with Statistics NZ Household Labour Force Survey (HLFS) data and links through to Statistics NZ business information in the prototype LBD.
- Longitudinal Business Database (LBD) prototype – based on a longitudinal business frame, with data from Statistics NZ business surveys or other outputs and external administrative data from other government agencies.

The replication of data and data processing systems is inefficient. The integrated datasets on individuals are illustrated in figure 2.

Figure 2



Also, these environments are inflexible; these integrated datasets are not able to efficiently handle frequent changes in administrative data made in response to policy or real-world changes. They create barriers to data integration due to their separate and independent approaches to storing data. To use existing and new data sources efficiently, we need to change our current siloed approach to linking data and create a new flexible data integration environment. In addition, redevelopment of our IT infrastructure provides the opportunity to build in better access controls.

There are two main initial streams of work.

1. Consolidation and redevelopment of current integrated datasets, including data from: LEED, SLA, EOTE, Ministry of Social Development benefit-related data, and HLFS and the prototype LBD.
2. Linking migrant and movement data (from Department of Labour (DoL)) to iLEED.

This work will:

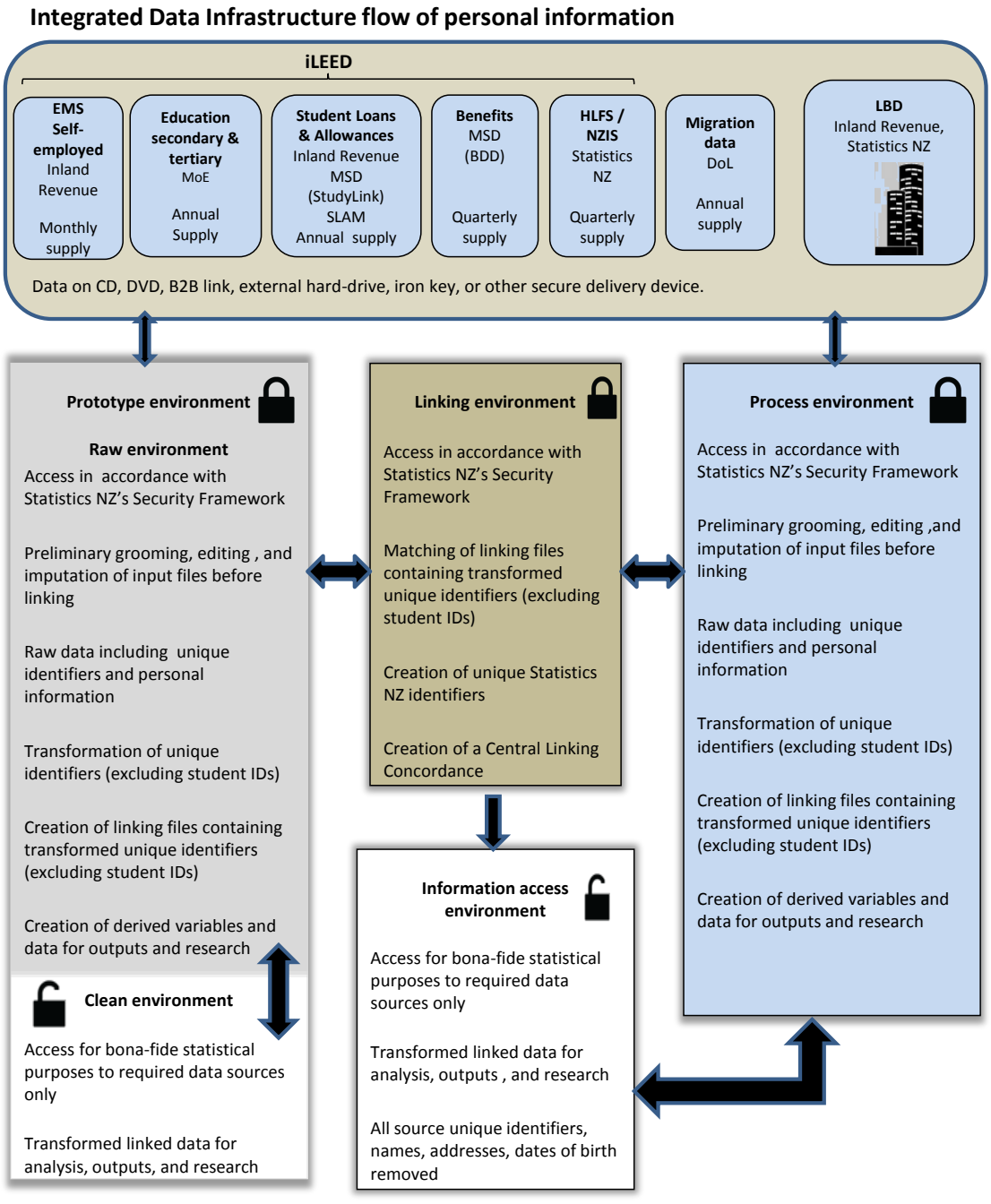
- create a statistical infrastructure that links administrative data from several government departments and Statistics NZ survey data. It will allow linking of individual and business-level data and provide a more systematic approach to longitudinal data linkage across the Official Statistics System.
- allow users to tell more detailed stories about people and businesses in New Zealand and investigate currently unanswerable questions.
- meet user demands for more powerful statistics than is currently possible from isolated datasets. This integration will occur without creating any extra burden on respondents.
- enhance the research opportunities available for developing and costing social and economic policy as well as evaluating government programmes.
- improve access to existing information sources, encouraging wider use and reuse of data.
- provide an environment for integrating further datasets in future in response to user demands for more or new information.
- maximise the use of existing administrative and survey data, and will:
    - reduce the time and cost of delivering new statistics
    - be able to respond to changes in existing data sources more efficiently.

- maintain appropriate security and access protocols – to ensure that data is protected and public confidence is maintained, only those with a bona-fide reason will have access to data.

2.2 Information flows

Statistics NZ will create a prototype environment within which to test, analyse, and develop the infrastructure, its processes and outputs. We will also create a full productionised infrastructure. The prototype will be replaced and enhanced by the full production environment.

Figure 3



Custom access to data across different sources will be created for individuals in accordance with Statistic NZ's Security Framework, and limited only to information that people need for official statistics or research work.

Individual unique identifiers will be retained in the prototype raw environment and the process environment. These will be transformed into new unique Statistics NZ identifiers in the prototype clean environment and the information access environment.

Files on CD, DVD, B2B link, external hard-drive, iron key, or other secure delivery device will be provided by the source data agencies listed below. Collecting data sourced from other agencies with their unique identifiers creates increased privacy risks as the data is transferred from one agency to another. To manage this risk, all agencies will encrypt their data in accordance with the recommendation from the Office of the Privacy Commissioner. A password (of sufficient strength) to unencrypt the data will be supplied. The data will be collected from the agencies by a Statistics NZ employee, or delivered by a representative of the agency. The timing of delivery may change. The passwords will not be written or carried with the data and no other stops will be made by the person carrying the data. Data mediums will be destroyed or returned in the same manner.

On arrival, the data will be copied to a ‘raw’ collection. Historical data from the source agencies is held permanently on this server. Iron keys or external hard-drives are returned to the data suppliers in person. A description of security measures for the data once it is held by Statistics NZ is described further in section 3.3.

It is when the individual records remain identified that the highest potential privacy risk exists. Any casual or intentional access to the data at this time might allow those viewing the data access to personal information. To mitigate this risk, all data will be stored in dedicated IT collections at Statistics NZ. The data is described in section 3.1.

Access to the data where unique identifiers, names, addresses, dates of birth, and education providers’ student ID numbers are retained will be to a small number of authorised Statistics NZ staff only, in accordance with Statistics NZ’s Security Framework. Authorised staff include:

- statistical analysts and custodians
- selected IT systems administrators, business analysts, and developers
- security auditors.

Security of the information will be maintained by limiting who can view the data to selected IT staff and a small team of statistical analysts (around 10) responsible for processing, cleaning, and linking data, and ensuring that the production system is running properly.

Unique identifiers are replaced before linking is undertaken (excluding student IDs supplied for the education providers). Once the data is linked, cleaned, and made ready for the clean prototype or information access environments, then names, addresses, and unique identifiers (including student IDs) will be removed or transformed securely in the clean prototype and information access datasets for analysis and research. Further, when researchers are seconded to Statistics NZ for research projects or provided with access through the Statistics NZ Data Lab, specific access to data will be created that includes only the data sources necessary for their research project.

Thus, while the system as a whole will have information about education, income, student loans and allowances, employment and benefits, migration, and businesses, if a researcher only requires information about education and employment history, that will be the only data they will be able to view.

3 Privacy analysis

This section describes the privacy issues associated with using personal information in the infrastructure. It looks at source data, integration, storage, and use.

3.1 Source data

The infrastructure will use data from Inland Revenue, MSD, the Ministry of Education (MoE), DoL, and Statistics NZ's HLFS and Business Frame (BF), along with business data from Statistics NZ's prototype LBD.

3.1.1 Statistics NZ's Household Labour Force Survey data

The HLFS data consists of variables related to individuals and their work and labour force status. Variables from the main survey and supplementary HLFS surveys include:

- name, sex, date of birth, personal address, employer's address, ethnic group
- household composition/relationships
- labour force and work variables – part-time / full-time status, labour force status, work for pay or profit, absent from work, more than one paid job, actual hours worked, usual hours worked, prefer to work more hours, occupation, industry, highest school qualification
- the New Zealand Income Survey (NZIS) for the June 2007, 2008, 2009, 2010, and 2011 quarters – a supplement to the HLFS, the NZIS produces a comprehensive range of income statistics
- the Survey of Working Life – a supplement to the HLFS in the March 2008 quarter, the survey provided information on changes in the employment conditions, working arrangements, and job quality of employed people
- the Survey of Dynamics and Motivation for Migration in New Zealand – a supplement to the March 2007 quarter of the HLFS, the survey's primary focus was to understand the drivers behind internal migration in New Zealand.

3.1.2 Inland Revenue data

Inland Revenue data covers two areas of interest.

Employment and earnings

The first is information about employment and earnings from pay as you earn (PAYE) tax and self-employed people's tax filings. This information comes as identified unit records with data that includes an individual's:

- name, sex, title (eg Mr, Mrs), date of birth, resident indicator (New Zealand or overseas), IRD number, address
- taxable earnings (plus earnings not liable, and lump sum indicator) for work performed, including social security payments taxed at source
- tax deductions (PAYE, withholding tax, family support tax credit, student loan indicator amount)
- employer and employment start and finish dates.

As part of the LEED project there is an agreement between Statistics NZ and Inland Revenue that governs the provision and management of data between the two agencies. This remains unchanged for the IDI since Inland Revenue will not be providing additional data to Statistics NZ. For further information see appendix A.

The risks to individual privacy arising from the infrastructure have been addressed, and can be managed to a low level. The potential risk of negative public perception, regardless of validity, about individual privacy issues arising from the infrastructure has been reduced, with credible responses available.

This analysis has identified potential risks to personal privacy, public perception, and media from the fact that personal information from multiple government agencies is coming together. This privacy impact assessment process, privacy protection procedures, and procedures for dealing with complaints are designed to address these risks.

Student loans and allowances

The second area of information that Inland Revenue supplies concerns people who received student loans and allowances. The department supplies information on student loans, debt, and income for each year from 1992 onwards, including:

- name, sex, title (eg Mr, Mrs), address, date of birth, education provider code, IRD number, and residency status
- loan registration start and end dates and reason for registration end, type of identification used when registering for a student loan (eg birth certificate, passport), and identification number
- principal transferred from StudyLink, interest transferred, interest eligible for write-off, loan balance, interest compounded, interest written off
- details of student loan repayments (including overdue repayments, capital write-off, capitalisation and penalties for the current period and overall, the current rate of repayment assessed, and the net sum of all credit/debit transfers and refunds)
- gross earnings (from each employer if more than one) including salary/wages, benefits, New Zealand Superannuation, and withholding payments, as well as tax paid, employer industry code, employer's IRD number, and employment stop date
- gross interest, gross dividend, estate/trust income, overseas income, partnership income, shareholder salary, rents, self-employed income, other income, allowances, total expenses claimed, taxable income, tax on taxable income, total rebates, family assistance entitlement.

3.1.3 Ministry of Education data

Data collected by several agencies is coordinated and supplied to Statistics NZ by MoE. These datasets sometimes include general information such as name, date of birth, sex, ethnicity, iwi (tribe), residency status, national student number (NSN),[3] and student ID. The datasets are:

- MoE data on enrolments, and completions of all provider-based tertiary qualifications
- MoE data on courses
- Tertiary Education Commission data on learners in industry training and modern apprenticeships
- New Zealand Qualifications Authority achievement data of industry training learners, for all learners who participated in industry training
- National student index data from MoE on verified names, date of birth, and sex.

[3] Statistics NZ applied for and was granted permission to use the NSN number (see appendix D: NSN PIA) for the EOTE feasibility study and Integrated Dataset on Student Loans and Allowances. This allows Statistics NZ to link transitions, school achievement, and targeted training data to tertiary education data, and to information on jobs, income, and earnings – by linking verified names, sex, and date of birth for each individual with the same fields in other datasets.

MoE also supplies information about secondary students directly, including:

- transitions data – for each formal secondary school student this includes:
  - general information – age, date of birth, sex, ethnicity, NSN
  - transitions information – school attended, school type, and years of school attendance
- school achievement data:
  - standards – identification number of standard, version number of standard, type of standard, reporting school, and result of standard taken
  - qualifications – name of qualification, year qualification achieved, level of qualification, version number of qualification, and unique identifier number of qualification
- targeted training data:
  - general information – sex, ethnicity, date of birth, NSN
  - targeted training information – training programme code, employment history code, and highest educational attainment of learner before entry to programme.

3.1.4 Ministry of Social Development data

MSD supplies information covering two areas of interest.

Student loan and allowance recipients

The first, supplied by StudyLink, is on student loan and/or allowance recipients for each year from 1999 onwards. The data items include:

- name, date of birth, sex, ethnicity (voluntary question), education provider(s), student IDs, IRD number, social welfare number, address details of study and postal location (including city and postcode), iwi (voluntary question), study start date and end date, residential status, and whether in prison
- type and amount of loan payments, sum of repayments and refunds, loan balance and interest transferred to Inland Revenue, loan balance and interest untransferred, and whether in prison
- type and amount of allowance paid, number of weeks of allowance, student's income, number of partners, number of studying partners, total partner income, each parent's income (if parent-income tested), and field of study (New Zealand Standard Classification of Education)
- type and amount of other payments, including the accommodation benefit, A and B bursary, Step Up scholarship, Bonded Merit scholarship, and top scholar award.

Income-tested benefits

The second area of information that MSD supplies is data related to income-tested benefits, from the Benefit Dynamics Dataset. The data items include:

- social welfare number, IRD number, date of birth, sex, and ethnicity of both the benefit recipient and partner included in the benefit.
- details about the benefit spells – including the event that lead to the spell, incapacity reason codes, benefit type, spell start and end dates, and the occupation code, last date worked, and last weekly earnings of the benefit applicant
- location of the district office through which the benefit was paid.

MSD also supplies data on participation in employment assistance. The data items include social welfare number, type of employment assistance, name of provider or employer, and participation start and end dates.

3.1.5 Student Loan Account Manager

Data from the Student Loan Account Manager (SLAM) relates to the period before StudyLink administered the Student Loans Scheme, when universities were responsible for collecting student loan applications and forwarding them to MoE and Inland Revenue. Both MoE and Inland Revenue supplied a SLAM dataset.

- IRD number, date of birth, sex, ethnicity, postcode of the student's residential address while studying and their permanent postcode, and the student's attendance type (ie full time/part time)
- the provider and programme of study the borrower is enrolled in, and student identification numbers
- details of student loan payments (including the amounts borrowed for fees, course-related costs, and living costs), interest and repayments, and the loan start and end date.

Inland Revenue's SLAM dataset includes:

- IRD number, name, date of birth, sex, ethnicity, and provider and student identification numbers.

3.1.6 Department of Labour data

Beginning with the development of the IDI, DoL will supply migration data.

- Movement data – name, sex, date of birth, passport number, and dates of entry and exit from New Zealand will be provided for New Zealand citizens, residents, and temporary visa holders.
- Visa application data – information collected during the visa application process including: name, sex, date of birth, nationality, visa type, policy criteria, and, where applicable, occupation, job location, or study institution of visa applicant.

3.1.7 Business Frame / LBD link

Individuals' employers in iLEED will be linked through to the Statistics NZ Business Frame and the LBD, which contains business information from both administrative and survey sources. This information is not at the individual person level. Datasets that are currently part of Statistics NZ's LBD are listed in appendix I.

3.2 Data integration

Integration enriches the individual datasets by linking them together. Integration is efficient because it adds value to existing datasets with little or no additional compliance burden to the population.

Data integration for this infrastructure will use a number of different variables to exact match, and to probabilistically match, unit records from the different datasets. Examples include:

- transformed IRD number
- transformed MoE national student number
- individual's provider student ID number
- transformed passport number

- first and last names
- date of birth
- sex
- ethnicity
- physical locations of home and business at the street level, as well as:
  - territorial authority
  - post code
  - region.

Statistics NZ staff will have sole responsibility for integrating source data for this infrastructure. At the time the data are linked, the Statistics NZ teams that manage the IDI will be working with identifiable individual records (eg for matching purposes: name, date of birth, sex, personal address, and employer's address will be the linking variables).

In keeping with Statistics NZ's data integration protocols, relevant documentation relating to IDI will be published on the Statistics NZ's website.

Statistics NZ also has a generic data integration policy page on its website.

Should concerns be raised, Statistics NZ and source agencies are well-placed to explain the processes used to protect personal privacy (through compliance with the department's data integration policy and the Privacy Act 1993) and the security and confidentiality measures in place to protect personal information once it has been collected. Statistics NZ will ensure that the data it holds is secure and can only be accessed by authorised personnel. There will be ongoing consultation between Statistics NZ and data-providing agencies.

This risk has been further managed by:

- consulting with the Office of the Privacy Commissioner and source agencies
- designing systems and processes that minimise the exposure of identified personal information within Statistics NZ
- Statistics NZ's successful record with security of the LEED-MSD, Student Loans and Allowances, EOTE, and HLFS-LEED datasets.

3.3 Data storage

The Privacy Act 1993 requires that all reasonable steps be taken to ensure that personal information held by an agency is protected against:

- loss
- unauthorised access, use, modification, or disclosure
- any other misuse.

Statistics NZ's standard security measures and protocols will govern the management of the data used in the infrastructure. Statistics NZ is required to comply with the confidentiality provisions of the Statistics Act 1975 and the Security in the Government Sector protocols.

Statistics NZ has well-established policies, procedures, and systems in place to ensure adequate measures of physical and electronic security, including:

- physical security systems controlling entry to premises and sections of premises to authorised people only

- visitors to Statistics NZ are subject to strict registration and supervision procedures; systems ensure their activities are confined to legitimate business
- access to data is in accordance with Statistics NZ's Security Framework
- access to Statistics NZ's IT systems requires a valid userid and password. A Statistics NZ security office actively audits and reviews security processes and addresses new and emerging threats.

Additional security arrangements for the infrastructure.

- All data collections and associated electronic workspaces will be secured (access will only be authorised for project personnel who need to access data for specific tasks, and to selected IT administrators who are required to maintain the IT system).
- Datasets will not be available to third parties.
- Regular audits of individuals able to access the dataset will be made.

Statistics NZ is confident that the policies outlined in this document, and the ‘culture of confidentiality’ that exists in the organisation, provide a high degree of protection for data held by Statistics NZ.

3.4 Data use

The infrastructure will:

- provide data to be used only for statistical purposes, which will include research undertaken by Statistics NZ employees, secondees, and bona-fide research undertaken in the Statistics NZ Data Lab
- not provide operational data to be used for administrative purposes, though the data could be used in an operational way.

These uses align with Statistics NZ's core functions as specified in the Statistics Act 1975. Any person viewing data with raw unique identifiers or unidentifiable individual data will do so for Statistics NZ purposes. They will be required to read, sign, and comply with Statistics NZ and Inland Revenue declarations of secrecy, which are binding for life. These declarations effectively place the same obligations, responsibilities, and potential sanctions on external agency staff who are seconded to Statistics NZ, or provided with access through the Statistics NZ Data Lab, as apply to Statistics NZ and Inland Revenue employees.

Individuals may have the following concerns about the use of their information.

- The information might be used in a manner that is detrimental to their personal circumstances.
- Information might be released that identifies them and aspects of their personal circumstances.
- Unrelated information might be collected about them in an ever-growing database for non-specific purposes (ie 'Big Brother').

Statistics NZ has addressed these concerns by ensuring:

- the data is used only for statistical purposes
- the data is anonymised as early as possible during processing
- no identifiable information will be released
- access to the data is subject to Statistics NZ and Inland Revenue protocols
- access is provided only in accordance with Statistics NZ's Security Framework.

3.4.1 Microdata access

Microdata access is currently managed through secondment arrangements or access to the Statistics NZ Data Lab. Researchers who want to access data outside the Statistics NZ Data Lab are required to be seconded to Statistics NZ. Statistics NZ will review access requirements and delivery on an ongoing basis.

Access to the microdata is restricted in accordance with the requirements of the Statistics Act 1975 and Statistics NZ's microdata access policy. In simple terms this means that access to microdata will be limited to bona fide statistical research and carried out in a way that protects against confidentiality breaches.

3.4.2 Release of results

Statistical information released from this infrastructure will be ongoing, and unrestricted beyond standard business protocols concerning confidentiality of data. Statistics NZ has a strong culture of confidentiality and all outputs are checked to ensure no identifiable data is released.

3.4.3 Data retention

Because of the research value of the integrated dataset, an active dataset will be retained until the benefits no longer outweigh privacy concerns and risks. Benefits versus risks will be determined through three-yearly reviews. If a review determines that the benefits no longer outweigh the risks, then the integrated dataset will be archived or destroyed.

4 Privacy risk assessment

The significant privacy risks to Statistics NZ that have been identified are summarised in the table below.

Privacy risk	Risk probability	Risk / impact	Risk mitigated by	Assessment of residual risk
Adverse public perception, or rejection of the legitimacy of the proposed use of this information, such as intentional or accidental misuse	Medium	High	• Adherence to Cabinet directive [CAB (97) M31/14] • Office of the Privacy Commissioner informed about the infrastructure and feedback addressed • Transparency about infrastructure objectives and processes • Consultation with key stakeholders • Compliance with the Statistics Act 1975, the Privacy Act 1993, and other relevant legislation • Access granted to data in accordance with Statistics NZ's Security Framework • Access auditable • Dataset archived or destroyed when risks outweigh benefits • All data checked before release	Low

Table continued next page

Table continued

Privacy risk	Risk probability	Risk / impact	Risk mitigated by	Assessment of residual risk
Use of the integrated data for other than statistical research purposes	Medium	High	• Access for specific genuine research purposes only, for approved projects and approved researchers • Access only to specific source datasets • Additional use of the data is subject to formal approval processes by Statistics NZ • Dataset archived or destroyed when risks outweigh benefits	Low
The risk of a breach of security or confidentiality, involving data related to individuals, while in transit or under the custodianship of Statistics NZ	Low	High	Standard Statistics NZ physical and data security practices, augmented by: • highest level of physical security of, for example, servers • person-to-person delivery of encrypted data • high-level encryption of data, including tight security of encryption key • granting access to data in accordance with Statistics NZ's Security Framework • auditable access • all outputs checked by independent staff to ensure confidentiality is preserved	Low

5 Additional privacy-enhancing responses

5.1 Governance mechanisms

The Government Statistician is required to make decisions on linking, access, and retention that are consistent with the requirements of the Statistics Act 1975.

5.2 Other privacy responses

The major privacy responses operating within Statistics NZ and the additional processes employed for the infrastructure have been outlined in this document. A privacy checklist that applies in a general sense to all Statistics NZ linked-administrative data is in appendix J.

6 Future compliance mechanisms

The IDI will be maintained with Statistics NZ's standard compliance processes. Access and system security will be maintained throughout.

7 Conclusions

Creating the IDI will enable statistical outputs on the transitions and outcomes of people for:

- secondary education
- tertiary education
- the labour market
- the benefit system
- movements in and out of New Zealand
- links to business data.

This information, and more frequent or further data included in the IDI, will lead to a better understanding of flows through the various systems and their associated outcomes. It will also allow for evidence-based policy analysis and modelling.

The infrastructure will also allow:

- increased flexibility to respond to changes and development in source-agency administrative datasets and Statistics NZ processes and outputs
- increased security
- evidence-based research to evaluate policy.

The risks to individual privacy arising from the infrastructure have been addressed, and can be managed to a low level. The potential risk of negative public perception, regardless of validity, about individual privacy issues arising from the infrastructure has been reduced, with credible responses available.

This analysis has identified potential risks to personal privacy, public perception, and media from the fact that personal information from multiple government agencies is coming together. This privacy impact assessment process, privacy protection procedures, and procedures for dealing with complaints are designed to address this.

Appendixes

A: Linked Employer-Employee Data Project: Privacy impact assessment

The Linked Employer-Employee Data Project: privacy impact assessment report is available from www.stats.govt.nz.

B: Linked Employer-Employee Data (LEED)-MSD Data Integration Project: Privacy impact assessment

The Linked Employer-Employee Data-MSD feasibility study privacy impact assessment is available from www.stats.govt.nz.

C: Integrated Dataset on Student Loans and Allowances: Privacy impact report

The Integrated dataset on student loans and allowances privacy impact report is available from www.stats.govt.nz.

D: Adding Statistics NZ as an 'authorised user' of the National Student Number: Privacy impact assessment

Privacy impact assessment: adding Statistics NZ as an 'authorised user' of the national student number is available from www.stats.govt.nz.

E: Employment Outcomes of Tertiary Education (EOTE) Feasibility Study: Privacy impact assessment

The Privacy impact assessment for the Employment Outcomes of Tertiary Education feasibility study is available from www.stats.govt.nz.

F: Feasibility Study for Linking Household Labour Force Survey with Linked Employer-Employee Data: Privacy impact assessment

The Privacy impact assessment for the feasibility study for linking HLFS data with LEED data is available from www.stats.govt.nz.

G: Data integration policy

Statistics New Zealand's data integration policy is available from www.stats.govt.nz.

H: Microdata access protocols

Statistics NZ's microdata access protocols are available from www.stats.govt.nz.

I: Statistics NZ’s prototype LBD

Statistics NZ's prototype Longitudinal Business Database (LBD) is an integrated database of all economically significant businesses in New Zealand. It is based on a longitudinal business frame, with data from Statistics NZ business surveys or other outputs and with external administrative data from other government agencies added. Established in 2007, the LBD has a number of components.

LBD components and their availability

Abbreviation	Component	Availability
LBF	Longitudinal Business Frame	From April 1999
BAI	Business Activity Indicator	From January 1992
AES	Annual Enterprise Survey	From 1999
BOS	Business Operations Survey	From 2005
R&D	Research & Development Survey	From 1996
BFS	Business Finance Survey	One-off 2004
INN	Innovation Survey	One-off 2003
BPS	Business Practices Survey	One-off 2001
MEUS	Manufacturing Energy Use Survey	One-off 2006
ITSS	International Trade in Services and Royalties Survey	From June 1996
APS	Agricultural Production Survey	From 2002
LEED	Linked Employer-Employee Database	From April 1999
IR10s	Financial accounts	From 1999
IR4s	Company tax returns	From 1999
Customs	Overseas merchandise trade data	From 1988
GAP	Government assistance programme	From 2000
IPONZ	Intellectual Property Office New Zealand data	From 1996
GST	Goods and services tax	From January 1992

J: Privacy checklist

Compliance with Statistics NZ's data integration protocols and generic processes will mitigate many of the risks identified in this checklist. The risks are described in their 'raw' form so how well they can be mitigated by generic processes and specific actions can be assessed.

Privacy checklist

Item	Response
• Have security procedures for the collection, transmission, storage, and disposal of personal information, and access to them, been documented?	Yes (see section 4). Further documentation will be done as part of the infrastructure.
• Are privacy controls in place for the infrastructure?	Yes. For example, collecting information that is not available from other sources, and keeping the information long-term – but subject to the strict access criteria set by the Statistics Act 1975.
• Have technological tools and system design techniques been considered, which may enhance both privacy and security?	Yes. For example, rigorous security procedures for personal information access, physical security and access controls, and IT security and access controls.
• Has there been an expert review of all the security risks and the reasonableness of countermeasures to secure the system against unauthorised or improper collection, access, modification, use, disclosure, and disposal?	The Statistics NZ Security Office regularly monitors compliance with security procedures.
• Have staff been trained in requirements for protecting personal information and are they aware of policies regarding breaches of security or confidentiality?	Yes. This includes training programmes and signing the Statistics NZ and Inland Revenue confidentiality agreements.
• Are there authorisation controls defining which staff may add, change, or delete information from records?	Yes – restricted access in accordance with Statistics NZ's Security Framework. These controls include standard relational-database management controls.
• Is the system designed so that access and data changes can be audited by date and user identification? Does it 'footprint' inspection of records and provide an audit trail?	Database auditing features allow this.

Table continued next page

Table continued

Item	Response
• Are user accounts, access rights, and security authorisations controlled and recorded by an accountable systems or records management process?	Yes, this is controlled by HR and help desk processes. In addition, access to the data must be approved by the data custodian. The data custodian for the infrastructure will be the Work, Knowledge, and Skills IDI project manager. Functionality exists to allow us to run reports to see who has access to the data.
• Are access rights only provided to users who actually require access for the stated purpose of collection or for consistent purposes? Is user access to personal information limited to that required to discharge the assigned functions?	Yes, requests are logged in the Statistics NZ help desk system.
• Are the security measures sufficient for the sensitivity of the information recorded?	Yes
• Are there contingency plans and mechanisms in place to identify security breaches or disclosures of personal information in error? Are there mechanisms in place to notify security breaches to relevant parties to enable them to mitigate collateral risks?	• The Statistics NZ Security Office regularly monitors for compliance with security policies, carries out audit functions, and investigates actual and potential security incidents. • The Statistics NZ Security Office has escalation procedures for managing and reporting security incidents.
• Are adequate ongoing resources budgeted for security upgrades, with performance indicators included in systems maintenance plans?	Yes
• What steps are to be taken to make the public aware of the infrastructure? Are individuals covered in the source datasets aware of the dataset's use?	• Key stakeholders have been informed of the creation of the IDI. • The privacy impact assessment for this infrastructure will be published on the Statistics NZ website.