Privacy impact assessment for the Integrated Data Service

November 2025

Approvals

General Manager, Data Operations & Infrastructure
Privacy Officer

Privacy and legal framework relevant to this privacy impact assessment (PIA)

The Privacy Act 2020 forms the basis of the risk analysis matrix used to assess the privacy implications of data integration and use in the Integrated Data Service.

The Data and Statistics Act 2022 governs Stats NZ and our primary purpose(s) for collecting, using, storing, and disclosing information.

Stats NZ, as a government organisation, may also be subject to other laws such as the Public Records Act 2005. Other laws may override the Privacy Act 2020.

Crown copyright ©
See Copyright and terms of use for our copyright, attribution, and liability statements.

Citation
Stats NZ (2024). *Privacy impact assessment for the Integrated Data Service.* Retrieved from www.stats.govt.nz.

ISBN 978-1-991307-93-4- (online)

Published in November 2025 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

Stats NZ's key privacy risks and recommendations 4
Purpose 6
Scope 6
Summary 6
Benefits of integrated data 11
Data integration flow 12
Protecting integrated data 14
Security measures 17
Other privacy considerations 18
Glossary 21
Risk assessment 25
Appendix 1: Risk matrix tables 29
Appendix 2: Data integration flow 30

Stats NZ's key privacy risks and recommendations

Risk	Risk description	Risk rating	Recommendation	Accepted/declined and justification for decline	Responsible	Status (includes notes and dependencies	Residual risk rating
Data security vulnerability	Holding significant volumes of identifiable data to build and maintain the IDI and LBD creates potential data security vulnerabilities.	Very high	A Certification and Accreditation (C&A) process for the IDI and its components has been completed. The C&A expires in 11/08/2027. Ensure the timely recertification of the C&A.	Accept	Integrated Data Production team and Security team	To be completed before 11/08/2027	Very low
Unwarranted data retention	Risk of retaining personal data longer than necessary without defined deletion process for data that is no longer in use or accessible. No clear data disposal mechanism.	Medium	Establish a clear exit strategy that includes timelines for data disposal. Contact the Data and Information Management (DAIM) Unit for guidance on disposal practices and requirements (informationmanagementenquiries@stats.govt.nz)	Accept	Integrated Data Production Team and Data Lab Team	Incomplete	Very low
			Develop protocols for long-term retention of raw identifiable data used to build the IDI and LBD in accordance with Data and Information Management Policy and associated procedures and guidelines.			Incomplete	

Data security vulnerability and/or breach	Accidental or malicious exposure of personal data	Medium	Review usage, information sharing agreements and confidentiality agreements and ensure third-party compliance with NZ privacy laws and Stats NZ's security protocols.	Accept	Data Lab Team and Legal Team	Incomplete	Very low
			Ensure ongoing monitoring and consistent application of the Five Safes Framework across the IDS ecosystem.		Integrated Data Production Team and Data Lab team	Ongoing	
			Conduct six-monthly review of Stats NZ staff access to databases and maintain an incident response plan.		Integrated Data Production Team and Data Lab Team	Incomplete	
			Ensure staff that work on production of the Integrated Data Service are onboarded effectively, including signing of Certificate of Confidentiality, abiding by Stats NZ Code of Conduct, and specific training for staff on what is acceptable and inappropriate behaviour when working with raw data			Ongoing	

This Privacy Impact Assessment (PIA) should be reviewed **at least every two years**, and more frequently if required due to changes in legislation, regulations, standards, or any updates to the systems, platforms, software, or operational processes that underpin the IDS. An earlier review should also be triggered if there are significant changes to the way the service operates or manages personal information.

Purpose

This Privacy Impact Assessment (PIA) provides a comprehensive and updated evaluation of the privacy and other risks associated with integrating data to deliver the Integrated Data Infrastructure (IDI), Longitudinal Business Database (LBD), and associated services under the Integrated Data Service (IDS). It does not cover other uses of integrated data by Stats NZ, such as those within the Integrated Statistical Data Services (ISDS) programme, which are out of scope for this assessment. It replaces the 2017 *Integrated Data Infrastructure: Overarching Privacy Assessment* (developed under the Privacy Act 1993). It aligns with the requirements of the Privacy Act 2020 and the development of new integrated data products.

This overarching PIA contains a high-level evaluation of the privacy risks and benefits associated with providing data integration services. When a new data collection is added to the IDS, a separate data collection-specific PIA and a corresponding risk mitigation table are published. These documents provide detailed assessments tailored to the specific dataset and should be read in conjunction with this overarching PIA.

This 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 2020 and the Data and Statistics Act 2022
- Assess if a proposed data integration is consistent with Stats NZ's Microdata Access Guide 2022.

Scope

This Privacy Impact Assessment (PIA) evaluates the privacy risks and safeguards associated with Stats NZ's IDS, which facilitates the secure and responsible integration of administrative and survey data into the Integrated Data Infrastructure (IDI) and Longitudinal Business Database (LBD).

Specifically, this PIA focuses on the transfer, storage, integration, access, governance, use, and disclosure of that data within the IDS. This PIA takes a general view, focusing on the privacy risks and safeguards common to all data handled by the IDS. This PIA does not focus on the privacy risks and safeguards specific to a data collection. Those are assessed separately in a data collection-specific supplementary PIA. As such, for a complete understanding of the privacy risks and safeguards, both the Overarching PIA and data collection-specific PIA must be read together.

Summary

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

The IDS is a special instance of data integration that has been approved, and is funded, by the New Zealand Government. It is built from and contains integrated data collections from public sector agencies and organisations in New Zealand. The IDS includes two core integrated data products: the IDI and the LBD. They hold de-identified microdata about people and households (IDI), and about businesses (LBD). The two data products are linked through tax data. If authorised under Part 5 of the Data and Statistics Act 2022, the IDI and LBD can be accessed via secure Data Lab infrastructure.

Researchers use the IDS to conduct cross-sector research that provides insight into our society and economy. The research can help answer questions about complex issues that affect New Zealanders.

The integration of data has some key features that could present privacy concerns without good stewardship. These include:

- The data integrated into the IDS is held long-term
- The IDI and the LBD integrate data from various aspects of individuals' lives to provide a comprehensive view for research and policy analysis
- The data collections typically contain data collected as people access services and benefits they are entitled to, rather than data collected solely or primarily for research purposes
- The data is collected from third parties (public sector agencies and organisations), not directly from individuals
- Stats NZ receives identifiable data from third-party data suppliers to create and maintain the IDI and LBD. Before the data is available in the Data Lab, it is de-identified to protect individual privacy. However, Stats NZ retains the identifying data separately, under strict access conditions, due to the underlying technology architecture of the IDI systems. This information is retained solely to enable Stats NZ to effectively produce, maintain, and enhance the IDI, for example, by adding further data to the IDI and conducting quality assurance.

Data integration offers significant public value. Specific benefits include:

- Supporting demand for integrated research data that enables more granular research and statistical analysis and deeper investigations. For example, exploring health outcomes influenced by economic, environmental, or lifestyle factors
- Enabling the creation of new statistics and research outputs without placing additional burden on respondents
- Helping social policy experts identify and respond to areas of high need or risk through more targeted interventions
- Facilitating cross-sector collaboration and understanding of how different government services interact. For example, between health and welfare systems to improve prevention and management of health conditions
- Improving access to data for both government and non-government researchers
- Providing longitudinal data that allows researchers to address previously unanswerable questions
- Promoting wider use and reuse of data across government, enhancing efficiency and innovation
- Building capability in data integration and analysis within Stats NZ and across the Official Statistics System through shared environments and collaborative practices.

When weighing up the risks and benefits of data integration, access, and disclosure, we are guided by the Information Privacy Principles of the Privacy Act 2020, the Data and Statistics Act 2022, the 5 Safe's Framework, Ngā Tikanga Paihere, and the Stats NZ's Microdata Access Guide 2022. This ensures that data is used to support informed, lawful, and ethically responsible decision-making.
Integrated Data Service overview

Statistics New Zealand (Stats NZ) collects personal information from a range of sources (data suppliers), including public sector agencies, (private) organisations, and individuals (for surveys administered by Stats NZ).

Stats NZ collects this personal information when it considers the information necessary or desirable for the purpose of the production of official statistics, research, or both.

In almost all cases, Stats NZ collects this personal information under the legal authority provided by the Data and Statistics Act 2022, specifically by agreement or request under a clause included in Part 3 of the Act. However, exceptions include information from the Courts, which must be collected under the legal authority of an Approved Information Sharing Agreement under the Privacy Act 2020.

After collection, personal information is then associated with a specific set of permitted uses and access conditions. This is dependent on the type of data supplier, the purpose of collection, and the legal authority under which it was collected. Stats NZ records these details in formal documents authorised by the data supplier and Stats NZ, and in operational systems, to ensure that use of that information within Stats NZ conforms to those permitted uses and access conditions.

One of the permitted uses of personal information is for the linking of data to form integrated sets. This brings us to the integration of data into IDS for the purpose of official statistics and research under Part 5, and the focus of this PIA.

Stats NZ are the kaitiaki (stewards) for two core integrated data products: Integrated Data Infrastructure (IDI) and Longitudinal Business Database (LBD).

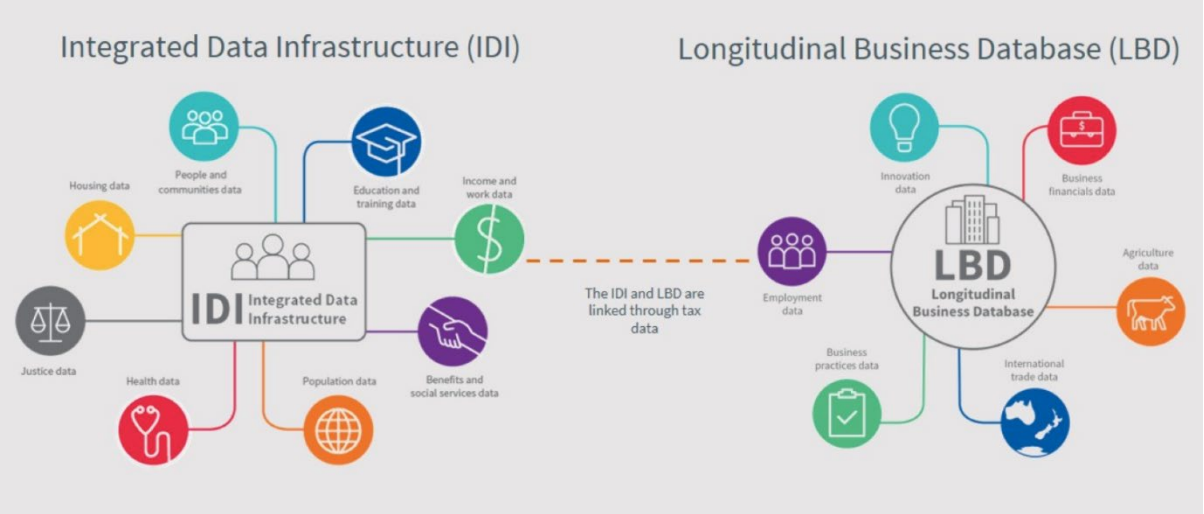
The information that Stats NZ collects includes personal identifiers, such as names, addresses, and dates of birth (DoB). Names are collected solely for the purpose of linking records across datasets and data collections. Addresses and DoBs are collected for multiple purposes. Within Stats NZ, access to identifiable data is strictly limited to a small group of highly trusted staff. This includes those responsible for producing and maintaining the IDI and LBD, linking specialists, and specified IT personnel, and a limited number of staff working on the Integrated Statistical Data Services (ISDS) programme. The ISDS programme is outside the scope of this PIA and is covered by a separate assessment

Once integrated (or linked together), this information is de-identified and made available in the IDI to approved researchers only. The IDI contains demographic information (such as age, gender, and ethnicity), and life event data (such as education, health, income, benefits, justice, and migration information). All identifiable information is removed or encrypted (replaced with another number).

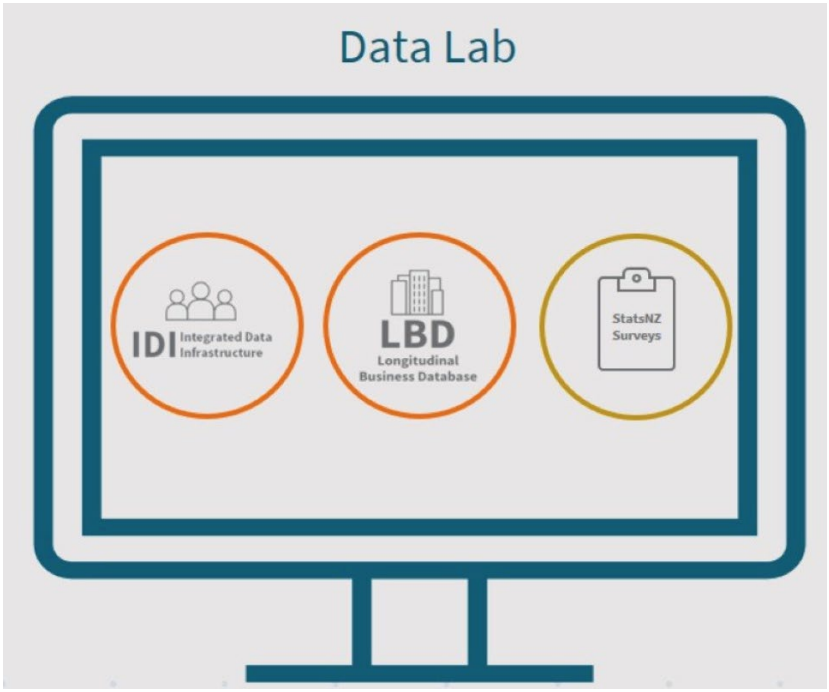
The LBD contains data about businesses, including financials, business practices, employment, and innovation. The data is collected within Stats NZ surveys and from public sector agencies. In the case of the LBD, information about sole traders is collected and treated as personal information under the Privacy Act 2020.

The IDI and the LBD hold de-identified information. If authorised under Part 5 of the Data and Statistics Act 2022, the IDI and LBD can be accessed via secure Data Lab infrastructure. The IDI and LBD are linked together through tax data, where IRD records can link an employee with their employer.

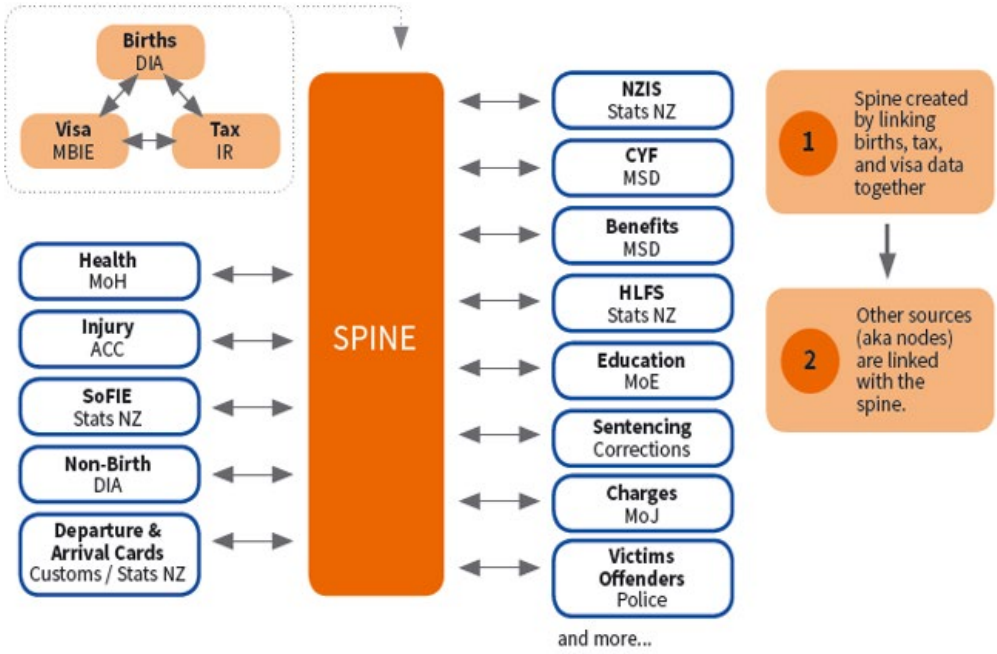
Researchers use the IDS to conduct cross-sector research that provides insight into our society and economy. The research can help answer questions about complex issues that affect New Zealanders.



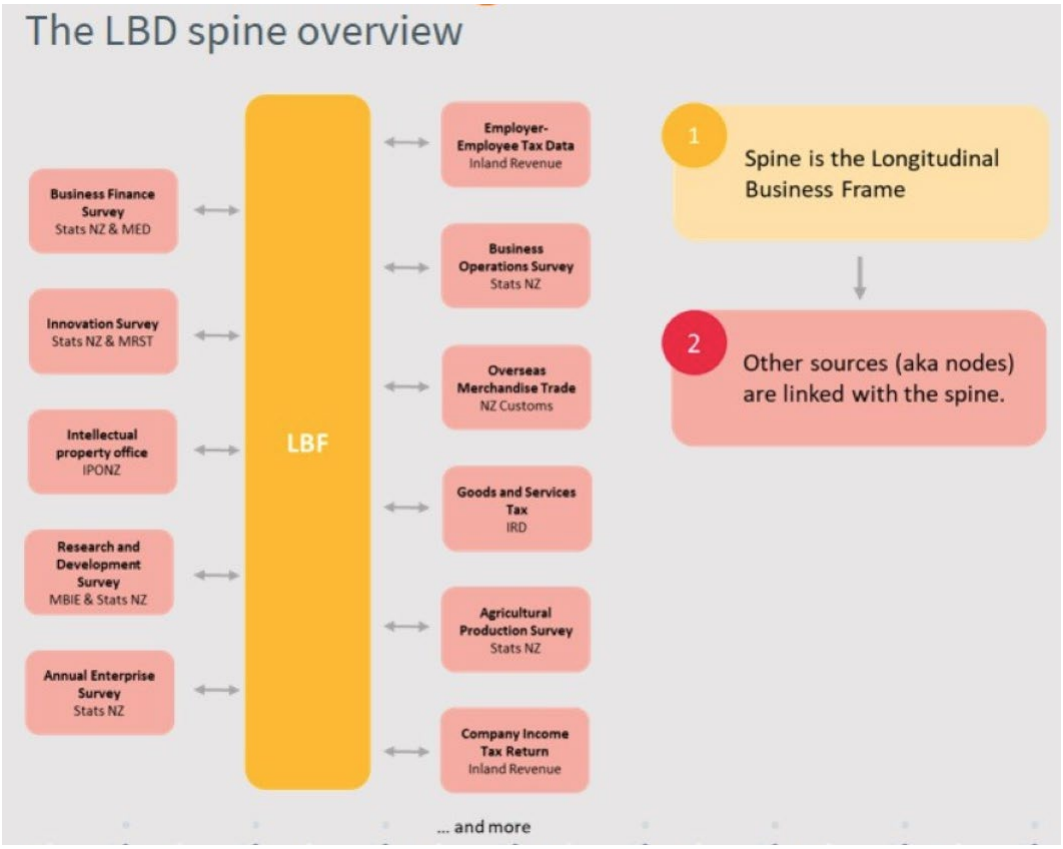
Stats NZ also manages and authorises access to **stand-alone data.** Stand-alone data is also held in the Data Lab's network attached storage (NAS) separately and is not linked to other databases. The stand-alone data largely consists of older Stats NZ surveys, but also includes surveys run by other public sector agencies, such as MSD's Youth Health and Wellbeing Survey. Since around 2021, there has been lower demand for using stand-alone data, as researchers find linked data (such as that within the IDI) more useful to understand the wide-ranging implications of policy and service provision, and how these affect outcomes.



The IDI is built around a central linking spine (a consolidated list of individuals), that enables consistent linkage across datasets. This spine is designed to represent New Zealand's 'ever-resident population' and is constructed using tax records, birth registrations and migration information. Additional datasets are then linked to this spine to ensure uniformity in data integration.



Similarly, the LBD uses a central spine formed from unit record data of businesses, including sole traders. This spine is created using information from Stats NZ Longitudinal Business Frame (LBF). Data on businesses' revenue, performance and productivity is linked to the spine using unique identifiers such as IRD number, New Zealand business number (NZBN), or Stats NZ assigned enterprise number. Researchers can also link LBD data with IDI data through encrypted employer IRD number found within the IDI's tax data, enabling cross-database analysis.



The Integrated Data Production (IDP) team updates (refreshes) the:

- IDI three times per year and refreshes are available at the end of March, June, and October each year
- LBD throughout the year, before taking and storing an annual snapshot of the database each December.

Benefits of integrated data

Informed decision-making for public good

Integrated data enables cross-sector research that provides insights into complex social and economic issues affecting New Zealanders. This supports evidence-based policy and service improvements.

Comprehensive view of society and economy

By linking data collected from public sector agencies, (private) organisations (NGOs), and individuals (Stats NZ surveys), integrated data offers a holistic view of life events such as education, health, income, and justice. This helps researchers understand patterns and outcomes at a population level.

Efficiency and reduced redundancy

Data integration and managed access, reduces duplication of effort across public sector agencies and organisations. It minimises the need for multiple public sector agencies and organisations to independently collect and integrate the same data, while efficiently managing access for all.

Enhanced data quality and consistency

Combining multiple datasets can improve the overall quality and reliability of information by enabling cross-checking and identifying discrepancies. However, under the Data and Statistics Act 2022, Stats NZ is not permitted to notify data suppliers of inconsistencies between datasets provided by other data suppliers (eg, differing addresses), limiting the ability to directly reconcile data across data suppliers.[1]

Privacy and security protections

Stats NZ has a strong reputation as data expert and trusted custodian of public information, and is bound by the Data and Statistics Act 2022 and the Privacy Act 2020 to protect the identities of people in the data it holds.

Stats NZ uses robust frameworks (like the Five Safes) to ensure integrated data is de-identified, only accessible to approved researchers for projects that benefit the public, and all research outputs are confidentialised before leaving the Data Lab. confidentialised before leaving the Data Lab.

Support for longitudinal analysis

Data integration enables longitudinal analysis by allowing researchers to observe patterns and outcomes over time in a de-identified manner. This is particularly valuable for understanding long-term trends, such as education and employment pathways.

1. Data and Statistics Act 2022, Part 3, Subpart 5, section 5.

Data integration flow[2]

Raw data is collected

Data is transferred by data suppliers into a secure file directory with Stats NZ's environment using secure file transfer protocols (SFTP) that provide end-to-end encryption and meet New Zealand Information Security Manual (NZISM) standards. Transfers can be completed manually via a web interface or automated from suppliers' systems (e.g. API connection). While IronKey may still be used in rare cases where SFTP access is not feasible, Stas NZ is transitioning to a secure enterprise Managed File Transfer (MFT) service for all standard data exchanges. The current system/platform has full and current security C&A.

Data is processed and linked

During the transfer, ingestion, and initial integration, the data remains identifiable, and access is strictly limited to a small group of highly trusted staff. This includes those responsible for producing and maintaining the IDI and the LBD; and a limited number of staff working on the spine for the ISDS. The IDP team, linking specialists and IT staff access incoming datasets via a secure file directory and load them into a secure, internal-access-only server. Personal information loaded for integrating into the IDI is stored in Microsoft SQL Server tables and is only used by authorised Stats NZ staff responsible for producing the IDI.

After cleaning the data containing personal information (such as names and dates of birth), the IDP team links the data. For some datasets, they also run a cross-referencing process to identify who might have more than one unique ID. This can happen in situations like bankruptcy, where a person may be issued a new IRD number, or in the justice system, where someone might be known by multiple names or aliases.

The goal of record linking is to connect two records from different datasets that belong to the same person or business. To do this, Stats NZ uses specialised software to match records accurately.

There are two main ways records are linked:

- **Deterministic linking** uses unique identifiers within datasets to make a direct match. For example, business records in the Longitudinal Business Database can be linked to Inland Revenue tax data using the employer's IRD number.
- **Probabilistic linking** uses personal identifiers within datasets to make a probabilistic match. For example, two records may be linked together if they have the same first name, birth date, address, and gender—but slightly different surnames—as they probably belong to the same person.

Probabilistic linking is not perfect. There is always a small chance of linking the wrong records (false positives) or missing valid matches (false negatives). For the purposes of official statistics and research, Stats NZ aims to keep false positives below 2 percent and estimates that false negatives are between 2–4%. This might be a higher tolerance of false positives or false negatives than you might expect from another public sector agency or organisation who need to make decisions about the services provided to an individual using the linked data. For their purposes, an incorrectly linked

2. Appendix 2 includes a visualisation and a detailed table outlining the technology used throughout the data integration process − from acquisition through to research output.

record could result in necessary services being denied or inappropriate services being provided. In extreme situations, this could be directly life-threatening.

The IDP team also conducts a de-duplication process. De-duplication is similar to cross-referencing, but it happens after the initial linking and covers a wider range of datasets. Its goal is to group together records that belong to the same person but appear multiple times. This process is used in projects where several records (called “nodes”) can be linked to one central record (called the “spine”). All matching node records are then given the same unique ID. Whether a project uses a one-to-one or many-to-one linking method depends on how likely it is that people have duplicate records in the dataset. This decision is made as part of Stats NZ’s methodology design. Detailed technical information about data integration, including de-duplication, can be found within this report: Data Integration manual: 2nd edition.

Data is de-identified

Before researchers can access data, it is de-identified.

Within the IDS this involves:

- removing the names of individuals and organisations
- removing individuals' day of birth and day of death, except when absolutely necessary and accompanied with additional scrutiny, such as for research into perinatal deaths or deaths very soon after birth
- assigning a unique identifier to an address, encrypting that unique identifier, and then replacing the street number, street name, suburb, and city with the encrypted unique identifier
- encrypting all unique identifiers assigned by data suppliers that are capable of directly or indirectly identifying an individual or organisation.

Unique identifiers include those assigned to those individuals and organisations, such as IRD, NHI, SWN or NSN identifiers. These have a direct and one-to-one relationship with the individual or organisation. For example, Inland Revenue assigns and manages each IRD identifier to ensure that each identifier is only ever used to identify a single individual and that a single individual is only ever identified by each identifier. Therefore, real-world knowledge of those unique identifiers would directly identify the individual or organisation. Encryption reduces this risk.

However, it also includes unique identifiers assigned to objects or events, such as a Police incident, court case, or lab test identifiers. These can have indirect one to one or one to many relationships with the individual or organisation. For example, one or many unique court cases might involve a single individual or organisation. Real-world knowledge of those unique identifiers would indirectly identify the individual or organisation. Again, encryption reduces this risk.

Stats NZ’s approach to encrypting unique identifiers in the context of integrated data or linked records also applies to publicly accessible information containing unique identifiers, such as the NZBN. With real-world knowledge of the relationship between a unique identifier and an individual or organisations, an unencrypted unique identifier would directly identify the individual or organisation. If this occurred within a linked record which contains an encrypted unique identifier that isn't publicly accessible, then the encrypted unique identifier would also now directly identify the individual or organisations.

Despite de-identification, if enough information about an individual and organisation is linked together, it can form a unique pattern. With enough real-world knowledge, an individual or

organisation could be identified. Attempting to identify individuals or organisations is explicitly forbidden. Stats NZ applies the Five Safes Framework to reduce that risk, which includes vetting of researchers and research projects before granting access to the data and the requirement that researchers attend confidentiality training.

Within the IDS, Stats NZ ensures the data is de-identified; before data leaves the IDS, Stats NZ ensures the data is confidentialised. Degrees of identification in data illustrates the difference between de-identification and confidentialisation is available.

De-identified data is made available for research

To access de-identified data, projects must request access to data for the purpose of official statistics or research and be demonstrably in the public interest. If access to that data for that purpose is permitted under the legal authority by which Stats NZ collected the data, then Stats NZ can authorise that access. Access is strictly limited to approved researchers working on approved projects, and they are only granted access to the specific data required for their research.

In some cases, additional access conditions may be applied to data at the time of collection. These conditions may be specified/requested by the data supplier and agreed to by Stats NZ. These conditions are typically designed to further safeguard individual privacy, protect commercially sensitive information, or mitigate risks associated with the integration and use of highly sensitive data.

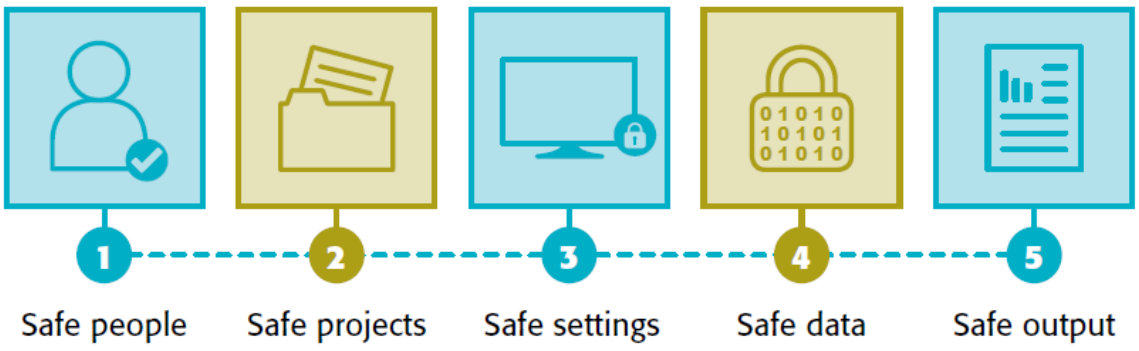
Where additional access conditions exist, Stats NZ records them in the operational systems that manage data access. These conditions are also documented in relevant documentation recording and formalising the data collection, such as a Data Sharing Agreement (DSA) or Memorandum of Understanding (MoU). When a project applies for access, the Data Lab team checks for any such conditions and assesses whether the project meets them, as part of the broader access approval process.

Protecting integrated data

Stats NZ has a robust set of standard procedures that we use to manage and protect the information we hold in the IDS. Access to personal information within the Data Lab environment is tightly governed by Stats NZ's privacy, security and ethical standards. Researchers and staff must meet strict requirements before they can interact with integrated microdata.

The 'Five safes' framework

Based on international standards this framework outlines requirements for sharing integrated data. Stats NZ only provides access to integrated data if all 'five safes' conditions are met.



Safe people – Before commencing work, researchers must:

- pass referee checks
- sign a Certificate of Confidentiality, which is a lifetime commitment under the Data and Statistics Act 2022
- Sign a researcher undertaking to follow Stats NZ rules and protocols and a Tax Secrecy Declaration
- Complete Data Lab induction and confidentiality training

Breaking our protocols may lead to researcher being banned, blacklisted, or prosecuted.

The Data Lab team conducts an annual review of all active Data Lab projects. This ensures that all researchers with project access match those approved as part of the project or variation process. Should an incorrect researcher have access, the Data Lab team will investigate, take appropriate remediation steps, and lodge a report following the Stats NZ security reporting process.

Stats NZ staff conducting research are treated and subject to the same vetting and access requirements as external researchers. Staff supporting the Data Lab such as IT personnel, also receive training to understand and comply with Data Lab conditions and restrictions.
Personal information about **researchers** who access the Data Lab is collected as part of the application and vetting process. This includes details provided in application forms, CVs, signed agreements, and reference checks. For researchers affiliated with universities or research organisations, Stats NZ may also verify contact details using publicly available sources such as university websites, if discrepancies are found in the application.

Safe projects – all research conducted in the Data Lab must be part of an approved project. Researchers undergo a rigorous application process to demonstrate their project serves a statistical purpose and is in the public interest. The IDI and LBD cannot be used for individual case management or decision-making about specific people or families.

Each application is assessed to ensure the data access requested is appropriate for the intended research and the project aligns with public interest objectives. Access to microdata in the Data Lab is strictly limited to vetted researchers working on approved projects, and all proposals are evaluated against Stats NZ's Microdata Access Guide 2022. These guidelines ensure that projects meet the criteria for public good, statistical validity, and ethical integrity, as outlined in **Part 5** of the Data and Statistics Act 2022.

Where a project needs access to an additional dataset, a project variation request must be submitted.

Projects are also reviewed by other teams within Stats NZ or elsewhere. These may be the legal team, a subject matter team such as the Household Economic Survey team, the Centre for Data Ethics or Innovation, or externally, such as the Health New Zealand's data enquiries team.

Safe settings – a range of settings for the physical and digital environment are designed to prevent unauthorised access to data. Data can only be accessed through a secure virtual environment known as the Data Lab, and only in research facilities approved by Stats NZ. Data Labs must meet requirements aligned to the Government's Protective Security Requirements (PSR), such as lockable doors that require physical keys or swipe card access. Labs are audited annually to ensure all requirements are still met.

The digital settings include the Data Lab computers having no connection to the internet or the organisation's network, and no USB ports or printing facilities, so data cannot be copied out. System use is audited, and automated alerts are sent to Stats NZ Security for any unusual activity.

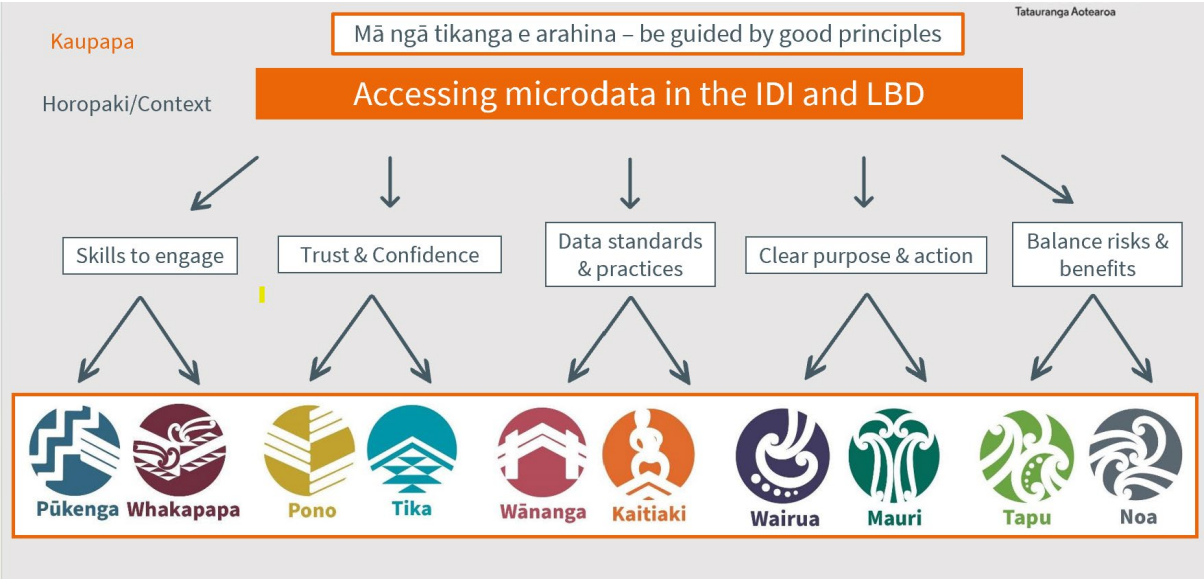
Safe data – data available to researchers is de-identified. This means information like names, dates of birth, and addresses has been removed. Numbers that can be used to identify people, like IRD and NHI numbers, are encrypted (replaced with another number).

Researchers are granted access to the specific datasets required for their approved project and only for the time agreed.

Safe output – researchers must ensure their work meets the requirements in the Microdata Output Guide 2020. Their outputs are reviewed and released to the researcher by the Data Labs team, once the team is satisfied all required rules to confidentialise data were followed. Where an output does not meet confidentiality requirements, the team will support the researcher to correct their work or will decline the output request.

Ngā Tikanga Paihere framework

This framework helps researchers use microdata in ways that are respectful, ethical, and culturally appropriate. While based on Māori concepts, it guides engagement and consultation with all communities in Aotearoa, including minority groups, people with disabilities, retirees, and others who may be affected by research using their data. The framework was designed to work in alignment and be complementary of the 'Five Safes'.



The framework draws in ten tikanga to promote ethical research guided by following good principles.

Pūkenga (knowledge and expertise) and Whakapapa (Community relationships) – Appropriate expertise, skills, and relationships with communities.

Researchers are not only technically and ethically prepared (Pūkenga) but are also accountable to, and guided by, the communities their research affects (Whakapapa).

This dual focus helps ensure research is both high-quality and culturally safe, with community voices genuinely shaping the research process and outcomes.

Pono (accountability and transparency) & Tika (value for all) – Maintain public confidence and trust to use data**.**

Research must be ethically grounded and impact driven. Researchers are expected not only to engage communities respectfully (*Pono*) but also to deliver value that aligns with those communities' aspirations (*Tika*).

This strengthens the legitimacy and relevance of research, especially when working with Māori and other communities whose data and lived experiences are central to the work.

Wānanga (organisations) & Kaitiaki (data stewardship and governance) – use good data standards and practices.

These principles support a research ecosystem where organisational structures (Wānanga) and community governance (Kaitiaki) work in tandem to uphold ethical standards and cultural integrity.

Organisations provide the frameworks and resources needed for responsible research, while communities provide the guidance and oversight to ensure data is used in ways that reflect their values and priorities.

This partnership strengthens trust, enhances research quality, and ensures that data use is both technically sound and socially responsible.

Wairua (community good) & Mauri (provenance and transformation) – Clear purpose and action.

Researchers must align their methods and objectives with community values (Wairua), while also being transparent and accountable in how data is transformed and used (Mauri).

This dual focus helps ensure that research is both ethically sound and socially responsible—delivering real benefits to communities, respecting their data, and minimizing risks.

By integrating these principles, researchers can build trust, foster genuine collaboration, and produce outcomes that are meaningful and respectful to all stakeholders.

Noa (benefit and opportunity) & Tapu (sensitivity and risk – Balancing risks and benefits.

These principles guide researchers to maximise the benefits of data for communities while rigorously protecting against risks and sensitivities, ensuring that research is both inpactuful and ethically sound.

Researchers are expected to be transparent and inclusive in sharing insights (Noa), but also careful and respectful in handling sensitive information (Tapu).

This balance fosters trust, ethical integrity, and community empowerment, ensuring that research is both beneficial and safe.

Security measures

Protective security requirements (PSR)

Stats NZ's manages integrated data in accordance with its standard security measures and protocols, and in compliance with the confidentiality provisions of the Data and Statistics Act 2022 and Protective Security Requirements.

When data is collected from organisations and public sector agencies under the Data and Statistics Act 2022, each data collection is documented and signed by Stats NZ and the data supplying

organisation or public sector agency. The document used to formalise and record the collection could be a contract, data request, data sharing agreement, or memorandum of understanding (MoU).

Researchers access the Data Lab environment (IDI, LBD and Stand-alone datasets) via Azure Virtual Desktop (AVD), either from Stats NZ labs, remote Data Labs, or through Stats NZ staff access. Within this environment, they can use SQL, SAS, Posit (R) and Stata, along with applications of the Microsoft Office Suite.

Stats NZ Data Labs are located in several approved facilities across New Zealand, including Stats NZ's offices in Auckland, Wellington and Christchurch. There are also approximately 40 remote, secure Data Labs located in organisations such as government agencies, tertiary education providers and private research organisations. Additionally, one remote Data Lab is located in Sydney, for work on a specific research project for the New Zealand Government.

All people accessing Data Labs must complete the required training and vetting processes prior to entry. Limited exceptions apply for essential personnel such as cleaners, tradesmen and first aid responders. However, where possible, Data Labs computers must not be used while these individuals are present.

All project applications contain an acknowledgement section with the commitment to provide copies of any research publications to Stats NZ. These publications are then uploaded into Storehouse where they can be viewed by any member of the public. The Data Lab team reminds researchers to provide any publications when projects are extended or completed.

PSR security classification information

Throughout the entire data lifecycle Stats NZ's employees, contractors and approved third parties are required to comply with the mandatory handling requirements under the New Zealand Government Security Classifcation Systems, Protective Security Requirements (PSR).

This commitment begins at the point of data collection, when Stats NZ and the data supplier establish or modify the data transfer process. As part of this set up, the data supplier must inform Stats NZ of the security classification of the data. Stats NZ will only proceed with data collection if the supporting technology, processes and policies are designed to meet the specified classification requirements.

IDS security certification and accreditation (C&A)

The length of the security C&A validity is determined by the level of residual risk. The IDS currently has an active C&A, most recently completed in August 2025, and is valid until 11 August 2027. A re-validation will be required upon expiry.

Other privacy considerations

Monitoring or recording of people's activities

The IDS does not use ingested data to monitor individuals. Researchers accessing data through the Data Lab work only with de-identified datasets and are permitted to release cohort-level insights (never individual-level information).

The Integrated Data Service (IDS) monitors system-level activity, including auditing staff and researcher access to ensure compliance with data governance and security protocols.

Stats NZ uses Azure Virtual Desktop (AVD) to identify inactive researchers. While their accounts remain active for potential future use, access to projects via AVD is removed through a semi-automated process managed by Stats NZ Technology Services.

The Data Lab team also manually reviews researcher activity. Those who haven't signed in for six months are switched to "Viewing Only" in Salesforce, and their Data Lab access is removed. Access can be reinstated upon request.

Researchers work in secure environments where printing, emailing, and internet access are restricted to prevent data leakage. System use is audited, and automated alerts are sent to Stats NZ Security for any unusual activity. Research outputs are only released after confidentiality checks by Stats NZ staff.

Access to servers and databases is managed through Active Directory (AD) groups, which are reviewed monthly by designated owners. Departing staff have their access automatically revoked through corporate offboarding processes.

All IDI and LBD servers are monitored using Microsoft Sentinel, with real-time threat detection and response provided by Kordia, Stats NZ's contracted Security Operations Centre.

Handling and storage of personal information

Personal information integrated through the IDS is stored in Microsoft SQL server tables and only used by Stats NZ staff responsible for the collection and linking of this data, such as the IDP team, linking specialists, and Technology Services staff, and stored on an internal access only server. The IDP team then encrypts and de-identifies data before transferring it to a separate SQL server in the Data Lab environment. After that, approved researchers can then access the datasets necessary to complete their research project.

Researchers' personal information is stored in Salesforce and Active Directory, to allow them access to the Data Lab environment to complete their research. While general researcher records are visible to Stats NZ Salesforce users, access to sensitive attachments is restricted to those with Data Lab Advisor or Approver profiles, which includes approximately 50 staff members. These profiles ensures that only authorised personnel can view confidential documents, maintaining separation between teams (e.g., Data Lab vs Collections). Researchers are made aware, via privacy statement, that Stats NZ collects their personal information, that they can request a copy of any personal information we hold about them, and to ask for it to be corrected if they think it is wrong.

Stats NZ staff personal information is used to manage access permissions for maintaining the IDI and LBD systems, including roles within IT and the IDI team. Access is primarily managed through Active Directory, which stores only staff names and usernames. In addition, the personal information of Data Lab team members will also be stored in Salesforce.

Code Modules monitoring

The team developing the Code Modules programme within the IDI and LBD uses system logs to understand how researchers engage with the modules (e.g. how many modules are used and how frequently). This monitoring is intended solely to support reporting on module uptake and inform improvements to the project's performance. As the Code Modules evolve, monitoring may be extended to better understand usage patterns, but will remain focused on project development and reporting, not on individual researcher behaviour.

User queries monitoring

Stats NZ retains audit logs of all activity on the SQL server used to access the IDI and LBD. Logs can be reviewed if misconduct is suspected. Spot audits are performed by running SELECT queries on SQL Server audit tables to identify access to sensitive datasets, unusual access times, or suspicious SQL commands, enabling targeted checks without continuous monitoring.

Information retention and disposal

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

All data managed in the Integrated Data Service 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. Disposal Authority 675 governs what data must be preserved, and what can be destroyed when no longer needed for statistical purposes.

Raw datasets in the Integrated Data Service are covered by class 7.2.1 of Disposal Authority 675. This class is for "Operational datasets" and includes "Data as received from all sources". The requirement is to retain the data until administrative and statistical purposes have ceased, after which it can be destroyed. Any research outputs can be kept until all quality checks and refresh purposes have been completed, then destroyed after release. The Data and Information Management Unit is contacted before destroying any data or information.

Any new collections and integrations of data into the Integrated Data Service will require their own PIAs, where more granular Retention and Disposal requirements will be applied.

IPP6 Privacy Act Requests

IPP6 Privacy Act requests for information that has been integrated into the IDI are not granted based on the rationale that Stats NZ does not hold the data in a way that makes the information readily retrievable. We encourage requestors to contact the public sector agency or organisations that supplied the data to Stats NZ and request the information from them directly.

However, individuals accessing the Data Lab can request copies of their personal information which is collected by Stats NZ to administer their access.

Glossary

Active Directory (AD)	Directory service developed by Microsoft for Windows domain networks. It acts as a centralised repository for managing users, computers, and other network resources, providing a hierarchical structure for organisation and access control.
Azure Virtual Desktop (AVD)	Cloud-based desktop and application virtualisation service provided by Microsoft that allows users secure access to the Data Lab environment from anywhere, using any device.
Certificate of Confidentiality (form)	Formal document signed by all researchers following their Data Lab induction. It outlines conditions of confidentiality, researcher's acceptance of these terms and states that intentional breaches may result in prosecution under the Data and Statistics Act 2022.
Clean databases	Processed and linked datasets ready for researcher use.
Confidentialisation	Statistical techniques used to prevent the disclosure of confidential information to unauthorised individuals in a way that could identify a person, household, or organisation. These methods offer a stronger level of protection than standard de-identification alone.
Data Integration	The process of combining data from multiple sources that were originally collected for different purposes.
Data Lab	Secure digital environment and a secure physical space.
Data Lab SQL server	Research-ready secure SQL server hosting cleaned and linked datasets accessible to researchers for analytics.
De-identify	The process where information that could directly identify a person or organisation, such as names, addresses, or passport numbers, is removed or encrypted.
Extract, transform, load (ETL)	Software application tools that automate the process of extracting data from various sources, transforming it into a usable format, and loading it into a target system like a data warehouse.
IBM QualityStage	IBM tool used for data cleansing, deduplication, and record linkage across datasets. The tool is

	integrated with Survey Portal for automated processes.
IDI Archive to CCL Vault V2	Secure archive platform for storing historical datasets and databases.
IDI Processing Manager	A module within the Survey Portal that manages the IDI processing pipelines.
IDI SQL server	Central server hosting multiple databases including IDI_RAW, Clean, LBD, and LEED, used for ingestion and processing.
IDI, LBD, LEED, Ad hoc	Different domains of integrated datasets used for research.
IDI_RAW database	Repository for raw ingested datasets before processing.
Information Linking Manager	A module within the Survey Portal that configures linking and matching rules for IBM QualityStage.
Integrated data infrastructure (IDI)	A large research database, holding de-identified microdata about people and households.
IR820 (form)	Inland Revenue form signed by researchers after completing their Data Lab induction. It must be completed by individuals who acquire, obtain or have access to revenue information. The form outlines that, knowingly failing to meet confidentiality obligations regarding such information, is an offence under sections 143D to 143EB of the Tax Administration Act 1994.
Longitudinal Business Database (LBD)	A large research database, holding de-identified microdata about businesses.
Longitudinal Business Frame (LBF)	A longitudinal register of all economically significant businesses in New Zealand, allowing for analysis of changes, trends and patterns over time.
Metadata	Data provided to describe and provide context for other data. This might include information such as source, structure, author, dates, variables.
Microdata (also referred to as Unit Record)	A data record relating to an individual, business, or entity. These records may include identifying information and therefore must not be released from the Data Lab in a form that allows identification.
Microdata Access Researcher Undertaking (form)	Form signed by all researchers after completing the Data Lab induction. It outlines the confidentiality obligations researchers agree to uphold. Breaching these undertakings may

	result in Stats NZ imposing measures, such as loss of access or potential prosecution.
Microsoft Managed Desktop (MMD)	Virtual desktop environment providing Stats NZ staff secure access to internal tools and resources. Deployed on a Surface Pros or laptops used by Stats NZ corporate staff to access internal systems.
Microsoft Office Suite	Licensed set of standard productivity tools (eg, Excel, Word) available to researchers.
Network Attached Storage (NAS)	A dedicated storage area connected to a network, providing file-based storage for multiple users and devices. It acts as a centralised storage solution for file sharing, data backup, and collaboration across a local area network.
Output checking	The process where data is released from the Data Lab, following reviews by the Data Labs team to ensure that no confidential information is released or disclosed.
Personal information	Information about an identifiable individual.
Posit (formerly RStudio)	Posit or Posit Workbench is a licensed suite of data science integrated development environment (IDE) tooling. R is the primary coding language. Posit is the licensed enterprise version of RStudio.
Refresh	An update of data in the IDI, where data is ingested from suppliers, linked together, and made available for research. At present, three refreshes of the IDI are conducted every year.
SAS / SAS Grid	A licenced statistical software used by staff and researchers for processing and analysis. SAS Grid refers to the distributed computing architecture implementation of the SAS product.
Secure directory	A controlled location within Stats NZ where raw data files are stored after transfer.
Security information and event management (SIEM)	Security solution that aggregates, analyses, and correlates security data from various sources within an IT environment to detect, investigate, and respond to potential threats in real-time. It combines security information management (SIM) and security event management (SEM) functionalities.
Security Operations Centre (SOC)	A third-party service that provides continuous monitoring, detection, and response to cybersecurity threats across an organisation's IT

	environment. The SOC operates 24/7, using tools like Security Information and Event Management (SIEM) systems.
Serv-U / Shift-It / IronKey	Secure File Transfer Protocol (SFTP) tools used for transferring raw data files from external agencies to Stats NZ.
Spine	A statistical spine is a central record, made up of variables or identifiers, used to link different datasets together over time. The spine of the IDI is made up of births, immigration, and tax data. These three categories pick up people born in New Zealand, people who migrate to New Zealand using a visa, and people who pay tax in New Zealand. The LBF is used as a spine for the LBD.
SQL server stored procedures	Scripts used to automate data loading and transformations. These are triggered via the Survey Portal.
Stand-alone data	Unlinked data, also available to use in the Data Lab. Includes Stats NZ surveys such as Household Economic Survey (HES), Household Income and Living Survey (HILS), General Social Survey (GSS), Household Labour Force Survey (HLFS).
Stata	Licenced data analysis and statistical software used in research. It is primarily used by researchers in the fields of economics, biomedicine, and political science.
Survey Portal	Stats NZ-developed web-based platform used to configure, schedule, and monitor ETL pipelines and metadata setup. Calls stored procedures for data processing.
WhereScape RED	Licensed ETL and orchestration tool used for automating data workflows, mainly for LBF and LEED datasets.

Risk assessment

Information privacy principle	Risks	If no or N/A, please explain why. If yes, please complete appropriate risk table section below	Recommendations from the Privacy Team/Mitigations	Rating
IPP 1 Purpose for collection: *lawful, connected with functions & activities, necessary for that purpose.*	Yes	There is a risk that the purpose of collecting data for the IDI may not be clearly understood or perceived as lawful or necessary by data subjects or stakeholders.	The collection is explicitly authorised under the Data and Statistics Act 2022, which provides for the use of data in official statistics and research enablement. The purpose is directly connected to Stats NZ's statutory functions and activities. The collection is necessary to support the IDI's role in enabling cross-sector research for public good. The overarching PIA and supporting documentation clearly articulate the lawful basis and necessity of the collection.	Very low
IPP 2 Source of information: *collected directly from the person it is about.*	Yes	Personal information is not collected directly from the individuals by Stats NZ which could lead to concerns about transparency or fairness.	The collection of personal information from third parties is authorised under the Data and Statistics Act 2022, which enables Stats NZ to collect data for official statistics and research. The use of indirect sources is necessary to support the IDI's function of integrating administrative and survey data for research purposes. Stats NZ publishes detailed information about the IDI, including its purpose, governance, and safeguards, to promote transparency. Privacy and ethical reviews are conducted to ensure that indirect collection is justified and proportionate. Data is de-identified before being made available for research, reducing privacy risks.	Very low
IPP 3 & IPP 3a What to tell an individual: *the organisation must be open about why they are collecting personal information and what they will do with it. This includes indirect (third-party) collection.*	Yes	There is a risk that individuals are not informed that their information may be shared with Stats NZ and used within the IDI for the production of official statistics, research, or both, as it is not collected directly from them. This lack of transparency may lead to individuals being unaware of or not anticipating how their information will be used, undermining informed disclosure and public trust in the data collection and use process.	The use of information for research purposes is authorised under Part 5 of the Data and Statistics Act 2022, and aligns with IPP2 exception g(ii). Exceptions in Principles 3(4)(e) & 3a (4)(g) apply. Stats NZ completes Data Sharing Agreements and PIAs for each new dataset added to the IDI, which are published on the Stats NZ website to support transparency. It is recommended that Stats NZ engage with suppliers to ensure their privacy policies, and where applicable, their terms and conditions, are updated to reflect the extended purpose of collection. This will help promote transparency for individuals. Privacy statements are maintained for researchers within the Data Lab and for the public on the Stats NZ website.	Very low

			Information about the Integrated Data Service and its purpose is publicly available, helping to inform individuals indirectly. Data is not published in a form that would identify individuals, reducing the risk of harm.	
IPP 4 Manner of collection: *personal information must not be collected by unlawful, unfair, or unreasonably intrusive means – the manner of collection must be fair and legal.*	Yes	There is a risk that individuals may perceive the collection of their personal information for the IDI as intrusive or unfair, particularly when not collected directly from them.	The manner of collection is lawful, fair, and not unreasonably intrusive, in accordance with the Data and Statistics Act 2022. Data Sharing Agreements and PIAs are completed for each new dataset added to the IDI and are published to support transparency and accountability. Information about researchers is collected solely to enable secure access to the Integrated Data Service, and researchers are required to sign documentation and read the relevant Privacy Statement to ensure informed participation. Public-facing information about the IDI and its governance is available on the Stats NZ website, supporting openness and public trust.	Very low
IPP 5 Storage and security: *safeguards must be in place to prevent loss, misuse, or disclosure of personal information.*	Yes	A Certification and Accreditation (C&A) process for the IDI and its components has been completed. The current C&A expires on 11/08/2027.	Timely recertification of the C&A of the IDI and its components is recommended to ensure a current C&A is always in place.	Very high
		Reviews are conducted to ensure research projects are using the correct data. Sometimes projects are given access to incorrect data, or researchers breach Data Lab policy (intentionally or accidentally). If this occurs, a Privacy and Security incident is raised. The researcher may be required to undergo additional training or be banned from further Data Lab activity in serious cases. Data is transferred and securely stored as per the normal process and settings for the Integrated Data Service. All Stats NZ staff, contractors and researchers must sign a Certificate of Confidentiality under s42 of the Data and Statistics Act 2022. Researchers must also sign a Researcher Undertaking and complete training in applying confidentiality methods.	Ensure staff that work on production of the Integrated Data Service are onboarded effectively, including signing of Certificate of Confidentiality, abiding by Stats NZ Code of Conduct, and specific training for staff on what is acceptable and inappropriate behaviour when working with raw data. Conduct six-monthly review of Stats NZ staff access to databases and maintain an incident response plan.	Medium
IPP 6 Access: *people have a right to ask for their personal information.*	No	IPP6 Privacy Act requests for information in the IDI are not granted based on the rationale that Stats NZ do not hold the data in a way that makes the information readily retrievable. We encourage requestors to contact the original agencies that collect the data (and supply data to the IDI), to request the data from them directly. Individuals accessing the Data Lab (researchers, etc) can request copies of their personal information in Stats NZ's system.	Not applicable	Not applicable

IPP 7 Correction: *a person has a right to ask an organisation to correct information about them if they think it is wrong.*	No	Under s29(2) of the Privacy Act 2022, Stats NZ is exempt for correction of data collected under the Data and Statistics Act 2022. Individuals accessing the Data Lab (researchers, etc) can request changes to their personal information in Stats NZ's system.	Not applicable	Not applicable
IPP 8 Accuracy: *an organisation must check before using or disclosing personal information that it is accurate, up to date, complete, relevant, and not misleading.*	Yes	Information is collected from data suppliers is considered 'as accurate, up to date, complete, relevant, and not misleading as we expect, based on agreements with data suppliers to provide necessary information about the quality limitations of the data. In instances where data does not meet expectations as agreed upon, Stats NZ investigates the issue to the extent where it has authority to do so, and this may result in resupplying or correcting the data.	Develop and document a standard process for identifying and resolving data quality issues when data does not meet expected standards.	Medium
IPP 9 Retention: *an organisation should not keep personal information for longer than it is required for the purpose it may lawfully be used.*	Yes	This information is specified as Stats NZ DA675, class 7.2.1., information is retained until "administrative and statistical purposes have ceased" and then destroyed. There is currently no timeframe or mechanism to delete data that is no longer in use or accessible.	Review general retention requirements, particularly where data is no longer used by a researcher (for data/datasets specifically requested for research under Part 5) and follow disposal practices and requirements and contact the Data and Information Management (DAIM) Unit for guidance (informationmanagementenquiries@stats.govt.nz). Observe Data and Information Management Policy and associated procedure.	Medium
IPP 10 Use: *organisations can generally only use personal information for the purpose it was collected.*	Yes	There is a risk that personal information may be used in ways not anticipated by individual.	Stats NZ uses personal information in accordance with its statutory role under the Data and Statistics Act 2022, specifically for enabling research and supporting public policy benefits. Use of personal information is subject to strict controls and safeguards detailed throughout the PIA, including governance, access restrictions, and de-identification. Personal information collected through researcher enrolment is used solely to manage access to the Data Lab and is not repurposed. Transparency is maintained through publication of PIAs and privacy statements on the Stats NZ website.	Very low
		Risk that data collected from suppliers for a limited purpose (research ***or*** official statistics), may be used for purposes beyond that originally stated.	Each dataset added to the IDI is supported by a Data Sharing Agreement and a separate PIA, ensuring that use is consistent with the original collection purpose and appropriately risk-assessed.	

			Stats NZ only approves projects that are clearly aligned with the stated purpose at the time of collection and ensures compliance through a robust project approval and ongoing monitoring process.	
IPP 11 Disclosure: *an organisation may only disclose personal information in limited circumstances.*	No	Data is shared by Stats NZ with researchers for the purpose of enabling research and the production of official statistics.		
IPP 12 Disclosure outside New Zealand: *rules around sending personal information outside New Zealand are contained in this Information Privacy Principle.*	No	All information in the IDI is hosted in Aotearoa/New Zealand, taking into consideration Māori data sovereignty. However, data may be accessed from the Sydney (Australia) data centre, meaning it will be disclosed to individuals and/or businesses overseas in these situations, raising concerns about cross-border data protection. Australia has similar and adequately comparable privacy laws and data protection measures to Aotearoa/New Zealand.	All information in the IDI is hosted in Aotearoa/New Zealand, ensuring primary data residency. In limited cases, data may be accessed from Australia, which constitutes overseas disclosure. Australia's privacy laws and data protection measures are considered comparable and adequate under New Zealand's privacy framework. Access is tightly controlled and limited to authorised individuals under secure conditions. The PIA outlines these scenarios and the safeguards in place to mitigate any associated risks. Stats NZ maintains transparency through public documentation and privacy statements.	
IPP 13 Unique Identifiers: *an organisation can only use a unique identifier when it is necessary. An organisation cannot assign a unique identifier to a person if that unique identifier has already been given to that person by another organisation*.	Yes	There is a risk that unique identifiers used to link records within the IDI could be disclosed to those with access to the data.	Unique identifiers are required to accurately link records within the IDI, but are removed or masked before data is made available to researchers. Stats NZ ensures that no identifying information, including unique identifiers, is accessible to those using the data. All IDI data is hosted in Aotearoa/New Zealand, and while access may occur via the Sydney data centre, Australia's privacy protections are considered comparable to New Zealand's. Access is strictly controlled, and only authorised users can interact with the data under secure conditions. These safeguards are documented in the PIA and supporting governance materials.	Very low

Appendix 1: Risk matrix tables

Risk ratings in this PIA are based on the information provided in the body of the PIA above. These may be further clarified or defined by relevant business groups based on other contextual information or work progress.

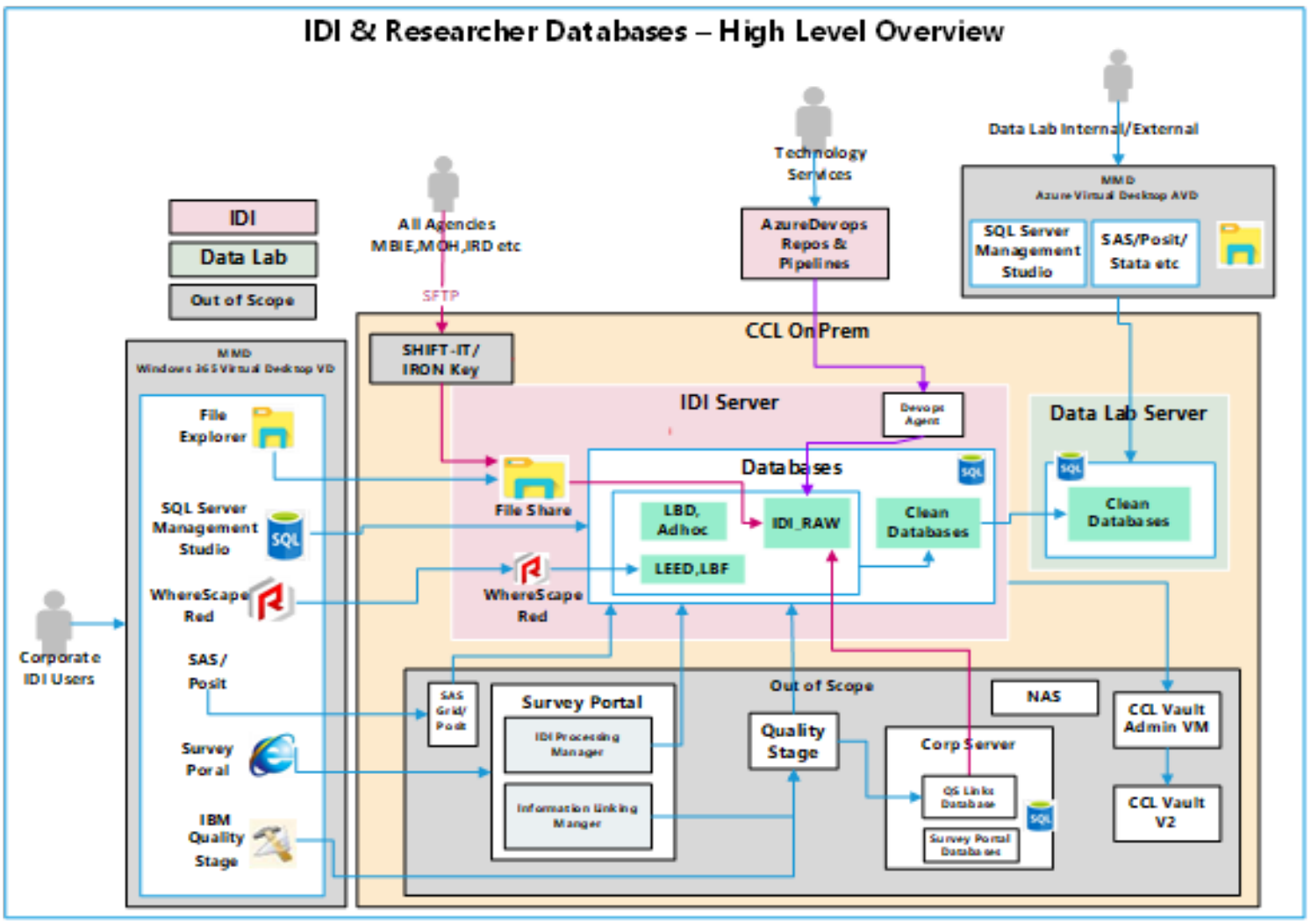
Risk rating matrix					
Likelihood	**Consequence**				
	Insignificant 1	**Minor 2**	**Moderate 3**	**Major 4**	**Severe 5**
Almost certain 5	Medium 5	High 10	High 15	Very high 20	Very high 25
Likely 4	Low 4	Medium 8	High 12	High 16	Very high 20
Possible 3	Low 3	Medium 6	Medium 9	High 12	High 15
Unlikely 2	Very low 2	Low 4	Medium 6	Medium 8	High 10
Rare 1	Very low 1	Very low 2	Low 3	Low 4	Medium 5

Risk likelihood	
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 likely 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

To establish risk rating, multiply likelihood by the consequence

Risk rating		
Very high	Score 20–25	Very high risk – stop activity immediately
High	Score 10–16	High risk – improve control measures and consider stopping activity
Medium	Score 5–9	Medium – review existing measures
Low	Score 3–4	Low risk – maintain existing measures
Very low	Score 1–2	Very low risk – maintain existing measures

Appendix 2: Data integration flow



This table describes each component of the data integration flow and who is responsible for it.

Data integration – components and responsibilities		
Component	**Description**	**Responsibility**
Data ingestion sources	All agencies (eg, MBIE, MOH, IRD) send raw data to Stats NZ. Data is delivered via SFTP, using secure transfer tools like Shift-IT or IronKey. Files are placed in a central File Share.	Data Operations, External Agencies
• **Processing environment**		
SQL – IDI	Central database server with multiple databases: IDI_RAW (raw ingested datasets), Clean Databases (processed and linked datasets), LBD, LEED, Adhoc (specialized datasets). WhereScape RED is used for ETL and orchestration of data movement and transformation for LBF and LEED datasets. IBM Quality Stage is used for linking and matching records.	Data Operations, ETL Management, Data Processing
IDI Processing Manager & Information Linking Manager	Part of the Survey Portal. Used to configure, schedule, and monitor processing pipelines. Allows metadata setup, source definitions, and linking configurations.	Data Operations, System Admins
Quality & integration	IBM QualityStage performs record linkage across datasets and phases. Survey Portal drives automated job runs for loading data into IDI_RAW, cleaning data, creating linked and derived datasets, and populating final output views.	Data Operations, ETL Management
• **Research-ready data**		
Data Lab SQL	Houses clean databases for secure researcher access. These databases are used for advanced analytics and cross-sector research projects.	Data Operations, Research Teams
IDI Archive to CCL Vault	Used for archiving databases and files.	Data Archiving, Data Operations
• **User access platforms**		
MMD (Surface Pros / laptops)	Used by Corporate IDI Users (Stats NZ staff). Provide access to File Explorer, SQL Server Management Studio, WhereScape Red, IBM Quality Stage, SAS, Posit, Survey Portal.	Internal users (Stats NZ Staff)
MMD/ThinClient – Azure Virtual Desktop (AVD)	Used by internal and external Data Lab researchers. Provides restricted, secure access to SQL Server Management Studio (SSMS), analytical tools (SAS, Posit, Stata), only connecting to Data Lab SQL clean datasets.	Internal & external researchers