Introduction to the Integrated Data Infrastructure (IDI)

June 2026

Social Investment Agency
Toi Hau Tāngata

New Zealand Government
Te Kāwanatanga o Aotearoa

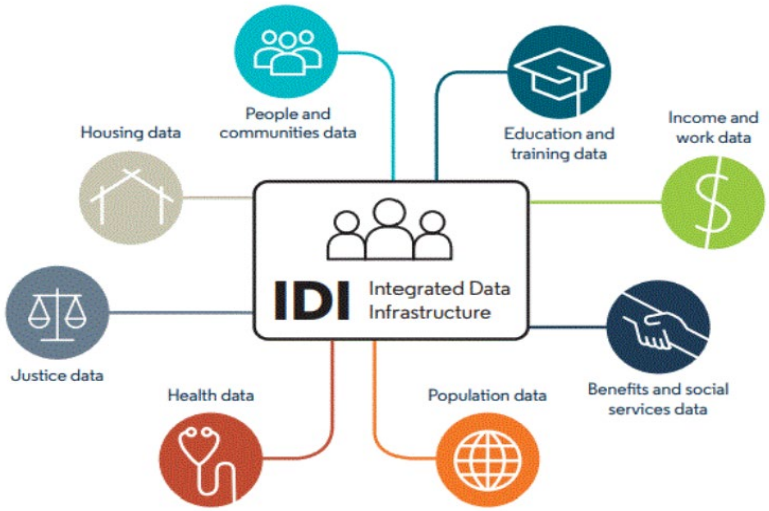
Stats NZ
Tatauranga Aotearoa

What is the IDI?

The Stats NZ Integrated Data Infrastructure (IDI) is a large collection of information (database). It holds de-identified data about people and households.

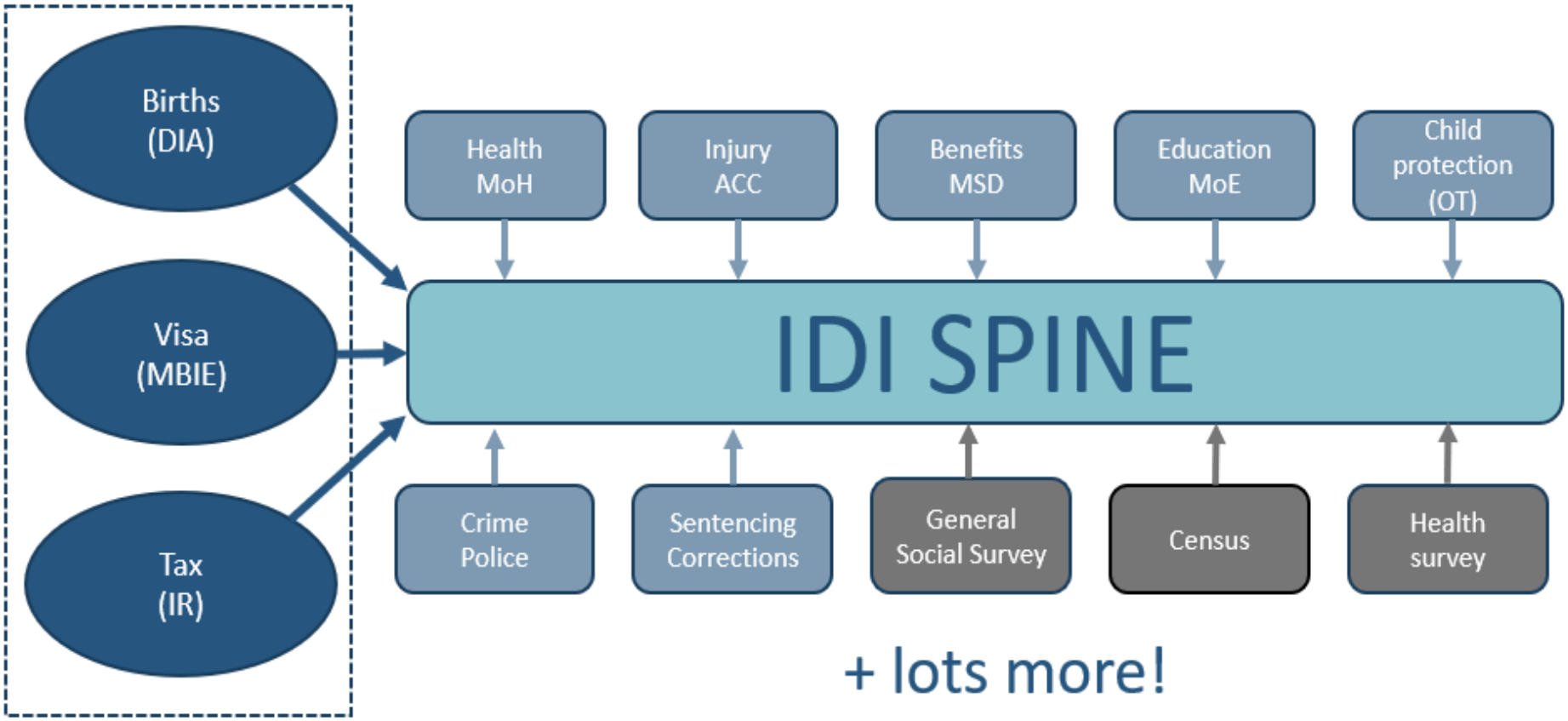
- The data is about life events, like education, income, benefits, migration, justice, and health.

- It comes from directly from individuals via Stats NZ surveys, and indirectly from public sector agencies and non-government organisations. The data is linked together, or integrated, to form the IDI.



What is the IDI?

The IDI Spine includes all New Zealand residents ever.



Why provide data to the IDI?

A key part of social investment is ensuring that organisations that provide social services – and the government – understand what works, and for whom.

- No matter how good your data collection is, you are not likely to have visibility of the wealth of data collected by government describing the lives and outcomes of the individuals and whānau you are working with.

- Using the Stats NZ Integrated Data Infrastructure (IDI) allows you to have visibility of this at a generalised (not individual) level, and where appropriate to learn from it and improve how you deliver services.

- In an outcomes contract, it also allows evaluation of progress against agreed outcomes.

Why provide data to the IDI?

For example: you know who you are working with – but across government, we also understand who you aren't.

It's possible to :

- Know how the people you're working with reflect your service delivery area (what percentage am I reaching? Am I demographically representative of my local area?)

- Understanding the activities and life events that your participants experienced before they started working with you

- Understanding the trajectories of your service users after they leave (e.g. did they stay out of the justice system? Hold stable jobs? Maintain good health?).

How is the data protected?

- The IDI does not contain people's names or personal details. The information has had names, exact birthdates, and addresses removed. ID numbers (like your IRD or health number) are encrypted so you can't be identified.

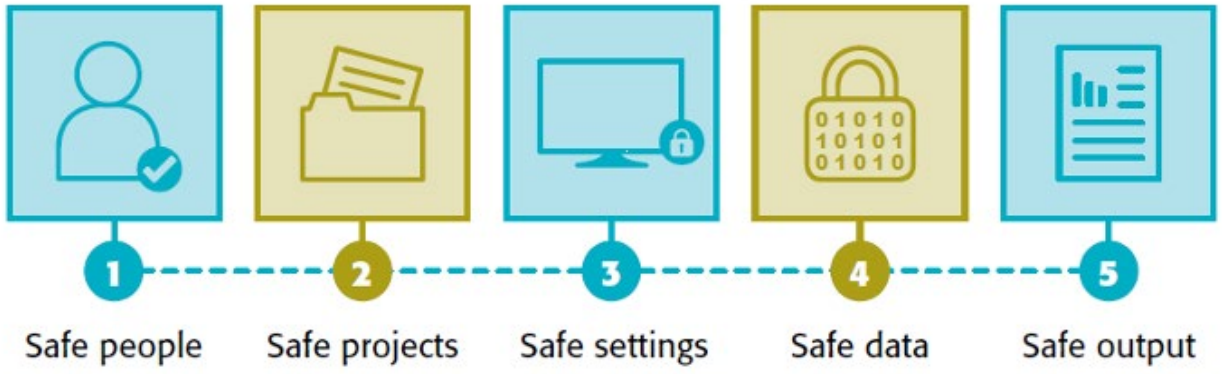
- You can't look up information about a specific person. Stats NZ can't search for or give you information about yourself or anyone else. The database is built for authorised researchers conducting research about **groups** of people **not** individuals.

- Stats NZ uses the '**Five Safes**' and '**Ngā Tikanga Paihere**' frameworks to manage safe access to the rich source of information about New Zealand people, households, and businesses available from the IDI.

What is the ‘5 Safes’ framework?

The 5 Safes framework is an internationally recognised, five-dimensional model used to manage risks when providing access to sensitive or confidential data, primarily for research.

It ensures data privacy by assessing Safe People (trained users), Safe Projects (public benefit), Safe Settings (secure IT/facilities), Safe Data (de-identified), and Safe Outputs (checked for disclosure).



What is the 'Nga Tikanga Paihere' Framework?

Ngā Tikanga Paihere is a framework intended to help researchers engage with Māori and other communities to ensure the use of microdata is respectful, ethical, and culturally appropriate.

Mā ngā tikanga e arahina - *Be guided by good principles*

Can individuals be identified within the IDI?

No. Data in the IDI is de-identified. This means information like names, dates of birth, and addresses have been removed so that no one individual can be identified.

- Importantly, even though data exists about de-identified individuals in the IDI environment, it can never leave that environment at the individual level.

- All data extracted from the IDI for analytics is confidentialised to ensure no individuals in the data can be identified. Stats NZ ensures that data is only used safely in ways that benefit New Zealanders.

What is de-identified data?

- **Identifiable data** contains personal identifiers such as names, exact dates of birth, and specific addresses.
- **De-identified data** has had personal identifiers removed or encrypted so that data records are not associated with named individuals.

For more information on how your data is de-identified in the IDI, please refer to the **Privacy Impact Assessment** (**PIA**) on the Stats NZ website.

How is data de-identified?

	Information supplied to Stats NZ for matching
Name	Bob Jones
NHI Number	123456
Date if Birth	16 Feb 1980
Passport Number	ABCDEF
Address	29 Acacia Road, Nutty town

	Information visible to IDI researchers
SNZ_UID	98989898
Birth Month	February
Birth Year	1980
Passport Number	7679203
Gender	Male
Meshblock	9999

	Information supplied to Stats NZ for matching
Name	Lisa Simpson
IRD Number	121212
Date or Birth	9 May 1981
NHI	8765432
Address	742 Evergreen Terrace, Springfield

	Information visible to IDI researchers
SNZ_UID	4545454545
Birth Month	May
Birth Year	1981
NHI	22222
Gender	Female
Meshblock	7777

What does data aggregation mean?

In the IDI, data aggregation means taking information about individual people, businesses, or events and combining it into summaries so that:

- No individual can be identified, and
- Researchers can only look at patterns or trends and not personal details.

What does rounding mean?

- **Random rounding** is a confidentiality rule Stats NZ uses to protect people's privacy when releasing results from the IDI. Data released from the IDI is randomly rounded to a base of 3

- **For example**: if the aggregated analysis in the IDI has a value of 31 the actual output from the IDI will be rounded up or down (33 or 30) so you cannot identify the precise number of records extracted.

What does suppression mean?

- **Suppression** is when Stats NZ does not allow certain numbers to be released at all because they could risk identifying a person or organisation.

- Data from the IDI that has a count of less than 6 will be **suppressed** i.e. **no** value will be contained in the output.

Who can access my data?

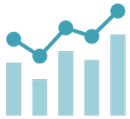
- After integrated data has had identifying information removed, only **vetted and approved researchers** can access selected, de-identified datasets for their specific project.

- This research **must** be for the **public good**. Users can only access the data in **secure** research facilities.

What the IDI is not

- It is **NOT** a complete picture of every New Zealander. The IDI only contains information that has been identified as **necessary** for research.

- It's **NOT** used to watch or track people or for public sector agencies to use when administering or making decisions about the provision of public services to individuals.



- The information isn't perfect or complete. For example, It might show the rates of disability across a group or the entire population during period of time, but it might not contain information about the exact type or nature of the disability.

Example of how IDI data is used

The Social Investment Agency's Regional Data Explorer supports access to data for regional and sub-regional decision-making.

This tool supports a social investment approach by helping to understand what population groups we need to invest in, and what works for these groups, and how we measure progress. Access to high quality data is an essential part of a social investment approach.