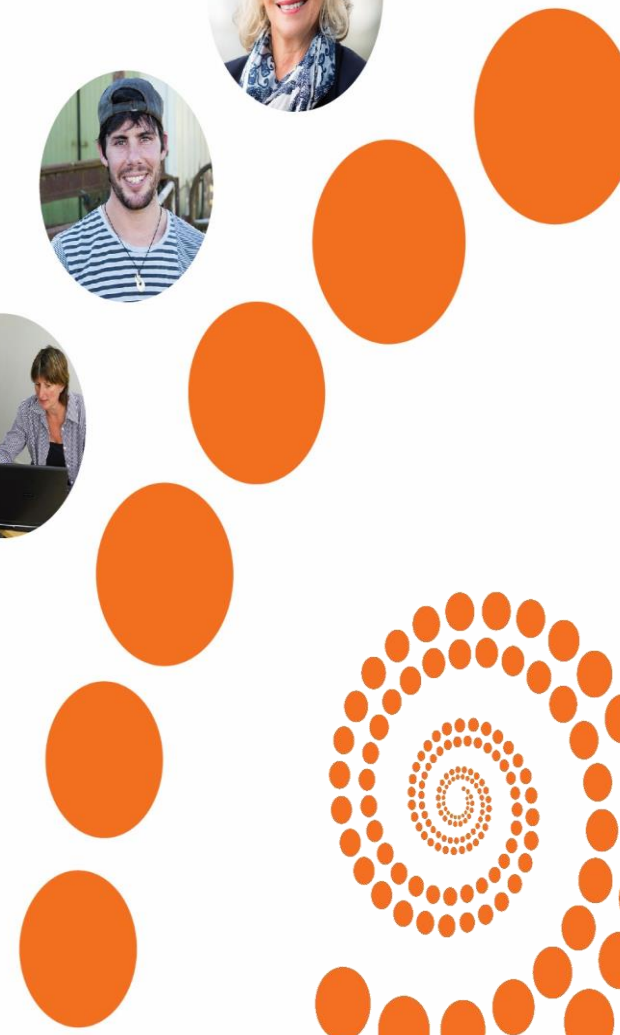# Integrated Data Infrastructure (IDI)

**Anna McDowell**

**Senior Manager, Integrated Data**
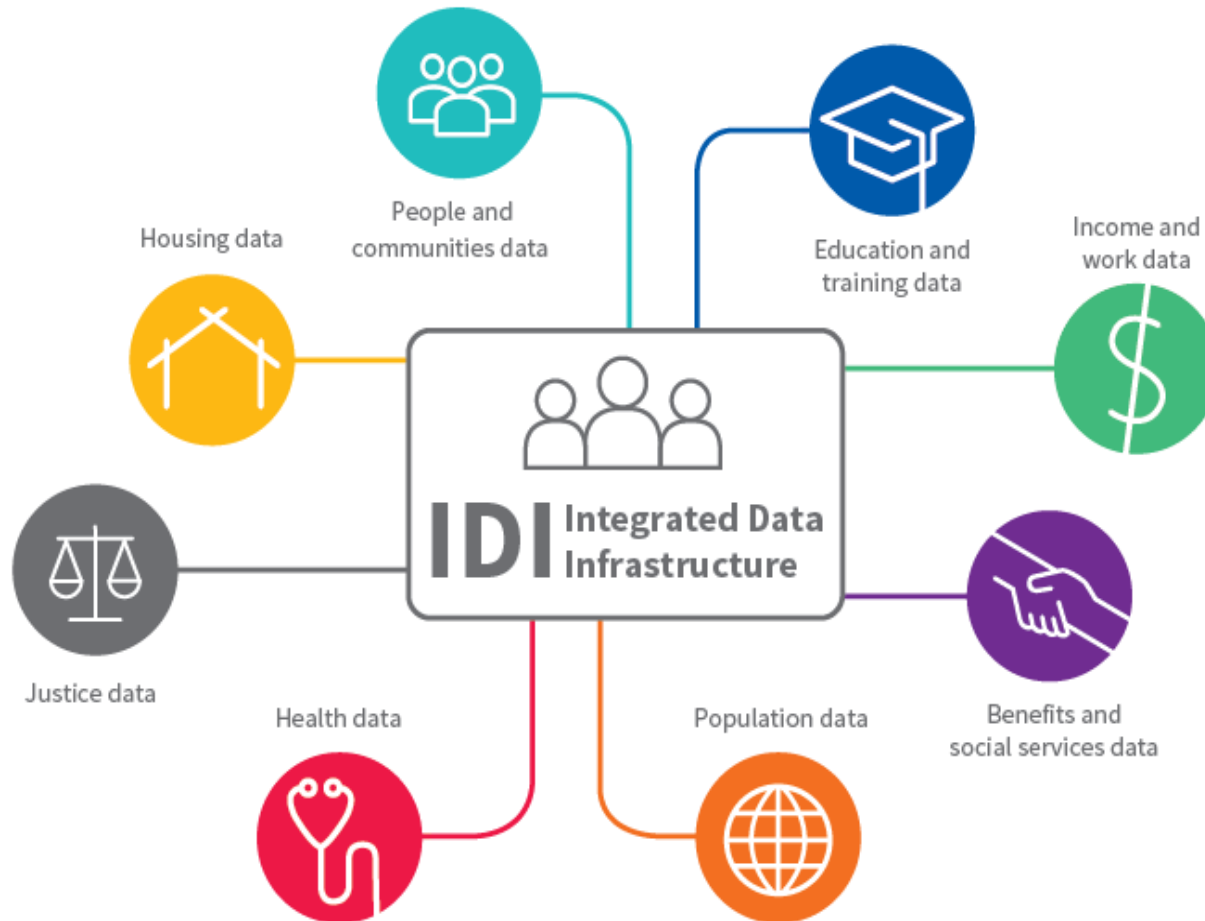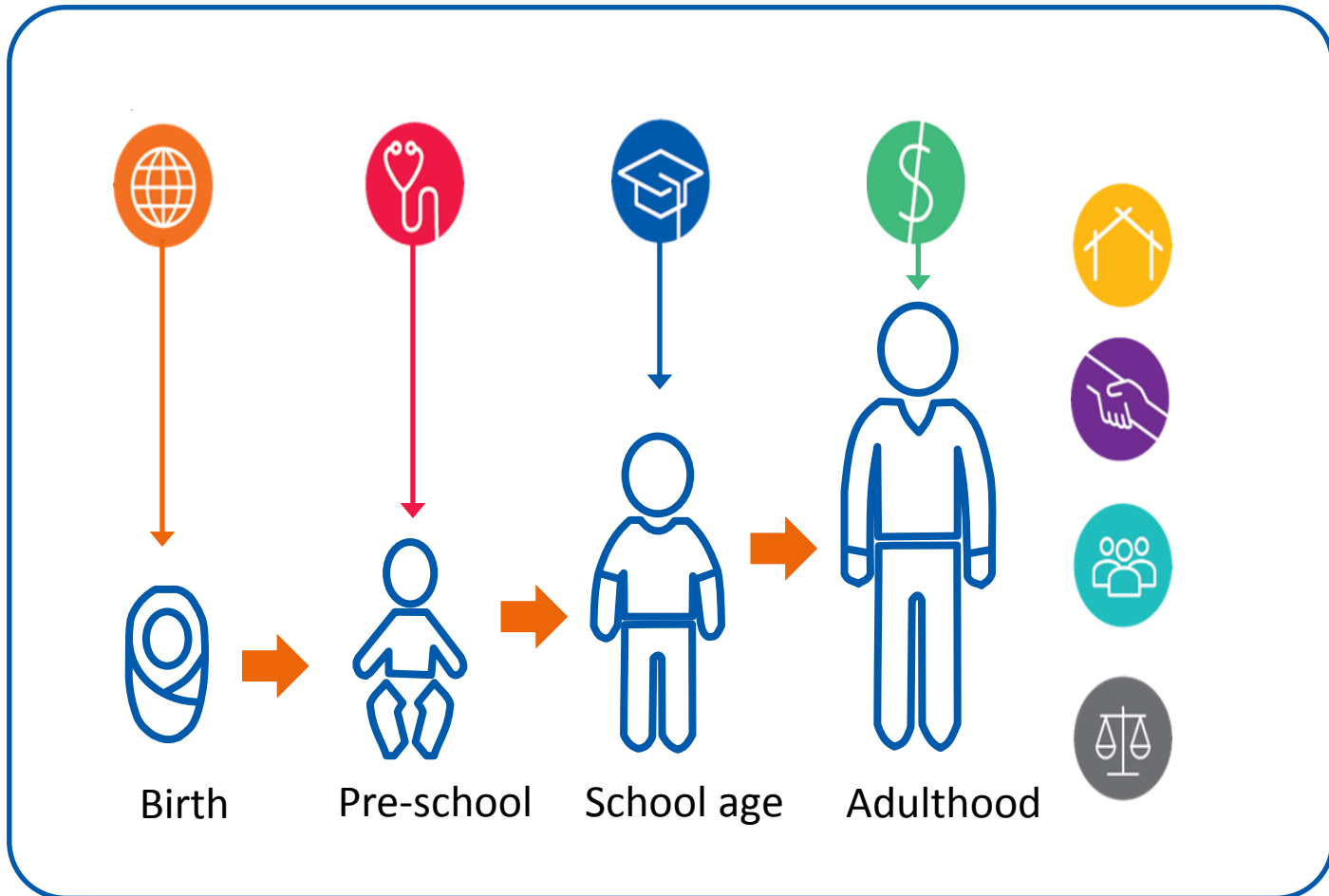
Oranga Tamariki, July 2018

# What is the IDI?

# Integrated Data Infrastructure (IDI)

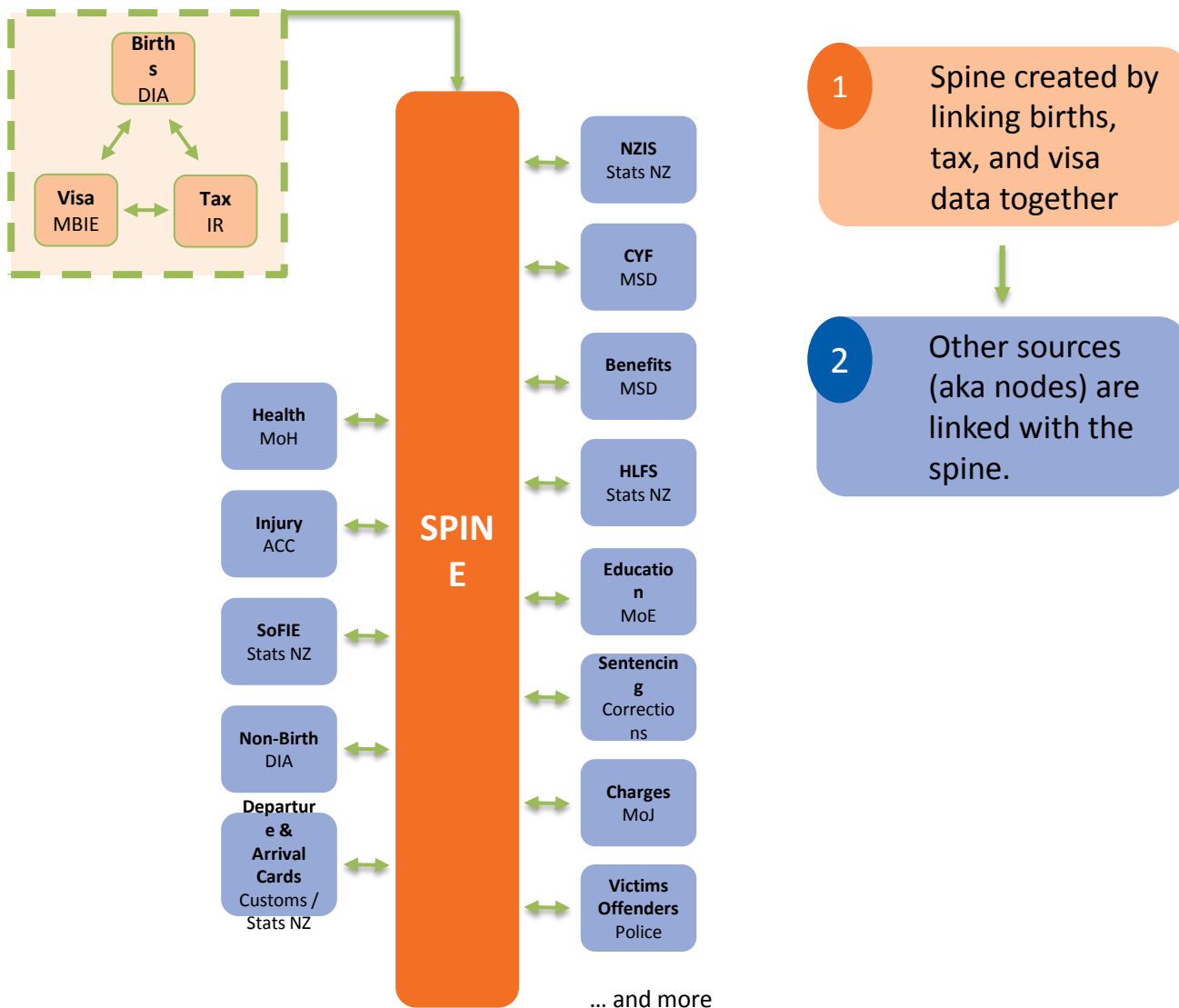An integrated database containing de-identified longitudinal microdata about people, households, & businesses.



For a full list of data, go to www.stats.govt.nz/idi

# A picture of people's lives ...



Birth    Pre-school    School age    Adulthood

# Linking datasets together



Births
DIA

Visa
MBIE

Tax
IR

SPINE

Health
MoH

Injury
ACC

SoFIE
Stats NZ

Non-Birth
DIA

Departure &
Arrival Cards
Customs /
Stats NZ

NZIS
Stats NZ

CYF
MSD

Benefits
MSD

HLFS
Stats NZ

Education
MoE

Sentencing
Corrections

Charges
MoJ

Victims
Offenders
Police

… and more

1 Spine created by linking births, tax, and visa data together

2 Other sources (aka nodes) are linked with the spine.

# The IDI's point of difference

*"We could still carry out our research with stand-alone datasets, but the IDI allows us to answer more questions. I could see people's change of location, but without linked data I can't see the events or information related to context of the move. We would have to make more assumptions, which we could still publish, but they would not have the same impact the IDI provides."*

Professor Gail Pacheco, NZ Work Research Institute, AUT, about her work on vulnerable transience.

# A world leader

### D5 – International Digital Showcase



> *"Internationally, only a handful of administrative datasets are on par with the IDI."*

**Professor Gail Pacheco, AUT.**

# How the IDI started

# History of the IDI

- 1997 NZ Cabinet endorsed Stats NZ for doing cross agency data integrations
- 2005 Linked employer-employee data (LEED) established
- 2011 IDI prototype established
- 2012 Government launched Better Public Services
- 2013 Cabinet agree to expand IDI under Analysis for Outcomes
- Growth and demand continues …

# Enabling factors

- Statistics Act 1975
- Stats NZ is a 'safe pair of hands'
- Ministerial champions
- Emphasis on data sharing and evidence based policy:
  - Better public services measures
  - Analysis for Outcomes
  - Social Investment Approach
- Social licence

# Privacy by design

# How is the data kept safe?

We operate within a 'five safes' framework to ensure that access to the IDI and LBD is only provided if all of the following conditions can be met:



| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Safe people | Safe projects | Safe settings | Safe data | Safe output |

# Access to the IDI for research

## Go

You need five greens before microdata access can be considered

- Research is for a statistical purpose
- Research is for the public good
- Research will be conducted by a credible team
- Suitable data is available
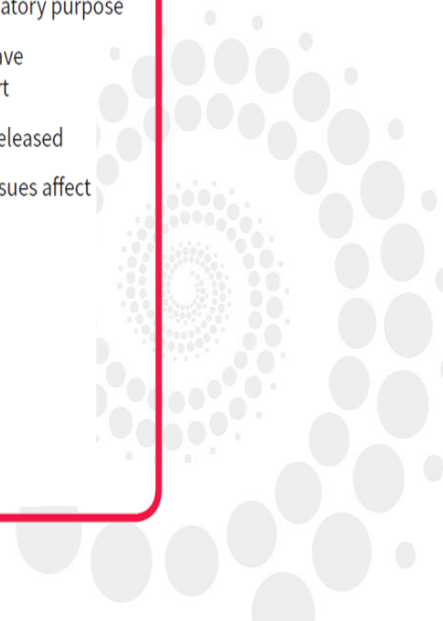- Stats NZ can enforce an agreement

## Check

You need to talk to us to check each orange issue for potential problems

- Research is for commercial gain
- Unrecognised research institution
- Data integration is required
- Researchers are based overseas
- Quality of data is not adequate for research purpose
- Lead researcher is a student or junior staff member
- Research is for service provider evaluation
- Non-government researchers are requesting access to business tax data

## Stop

You need to talk to us to resolve all red issues before microdata access can be considered

- Research is about named people or businesses
- Research is for a regulatory purpose
- Researchers do not have organisational support
- Research will not be released
- Complex legislative issues affect access

# What does the NZ public think?

# Social licence

"Social licence is societal acceptance that a practice that lies outside general norms may be performed by a certain agent, on certain terms. It is the result of a process of negotiation with a wider societal group, and means that the practice can be performed by that agent without incurring social sanction."
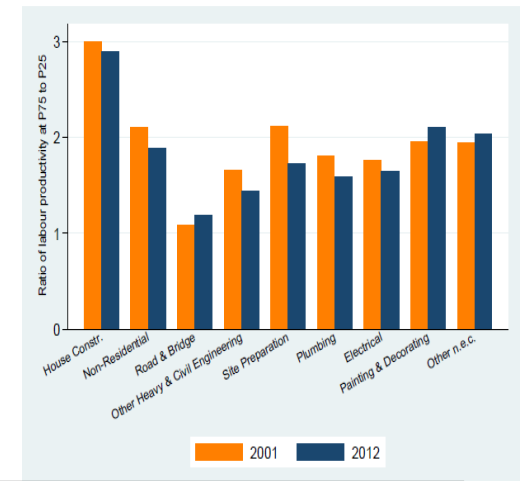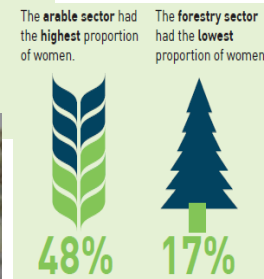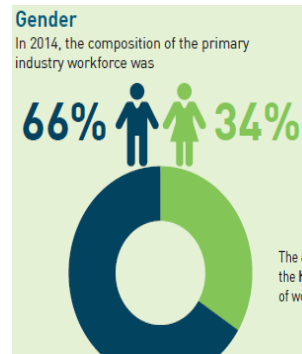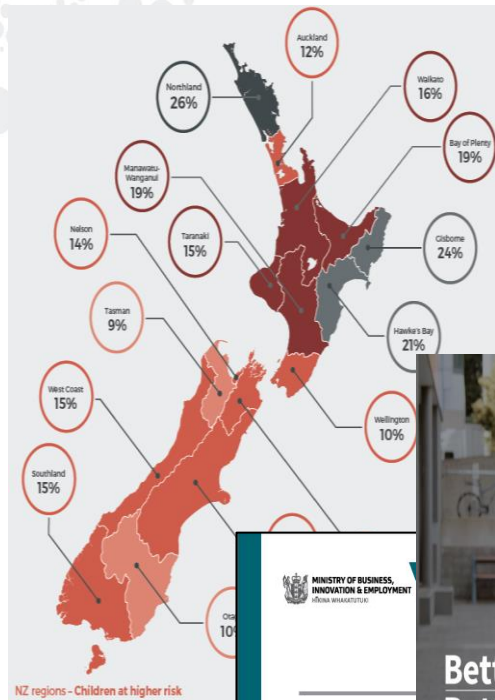
# Uses of the IDI

# How the IDI is being used
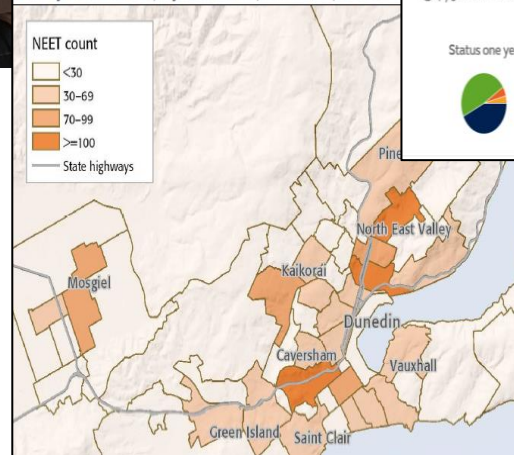
# IDI analysis

- Three types of analysis have been most important so far:

- evaluations and quasi-experiments
    - o propensity score matching is particularly widely used
    - o as well as other straightforward regression models

- micro-simulations
    - o combining many regression models and creating simulations with results

- profiling of sub-populations
    - o cross-tabs, counts and averages

# Outcomes of tertiary study

## Compare Study Options

Compare earning and employment information for different study areas.

**Compare Study Options**

Bachelors ▼

in Performing Arts ▼

with

Bachelors ▼

in Pharmacy ▼

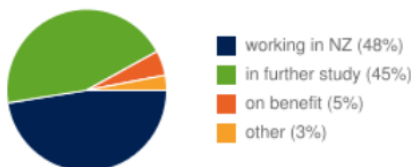**Compare**

What should I study?

### Bachelors: Performing Arts

$23,423  $26,823  $37,229

Median salary two years after study:

## $26,823

1 year after study   2 years after study   5 years after study

**60%** Employment rate two years after study

Status one year after study

- working in NZ (48%)
- in further study (45%)
- on benefit (5%)
- other (3%)

### Bachelors: Pharmacy

$38,600  $63,645  $72,963

Median salary two years after study:

## $63,645

1 year after study   2 years after study   5 years after study

**88%** Employment rate two years after study

Status one year after study

- working in NZ (97%)
- in further study (2%)
- on benefit (0%)
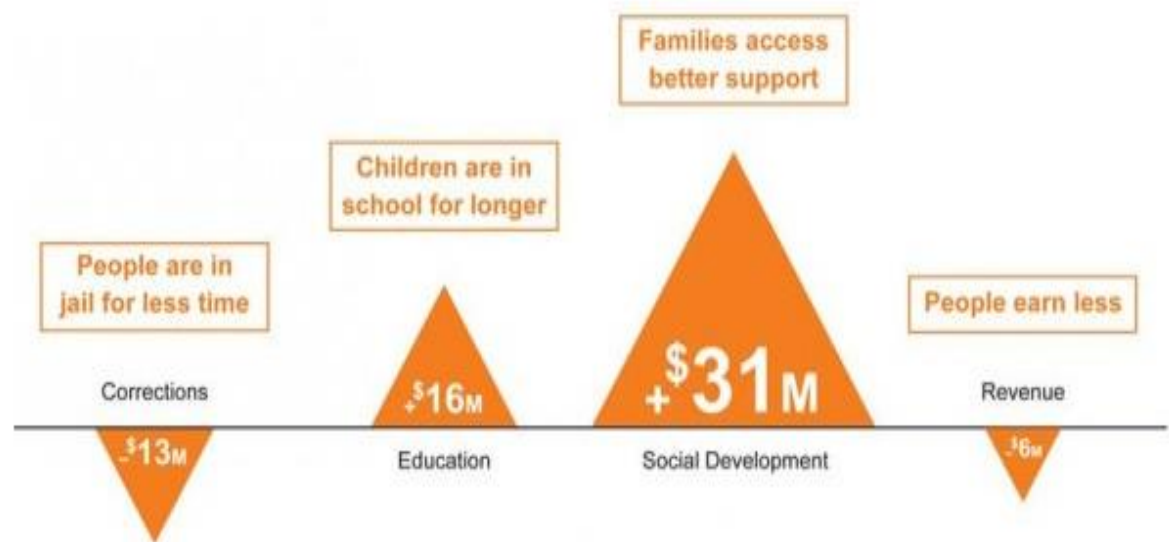- other (2%)

http://www.careers.govt.nz/ tools/compare-study-options/
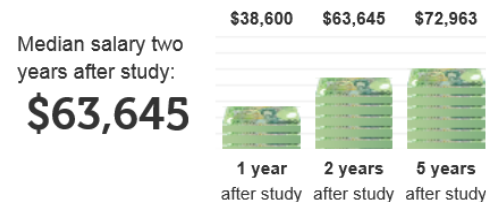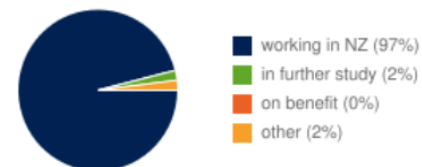
# Investing for social wellbeing

What is the impact of social housing on people's lives?



Note: There were no statistically significant findings for ACC, Health and CYF.

SIA also create reusable tools from their analysis - information about the Social Investment Analytical Layer (SIAL) and the Social Investment Measurement Map (SIMM) is online.
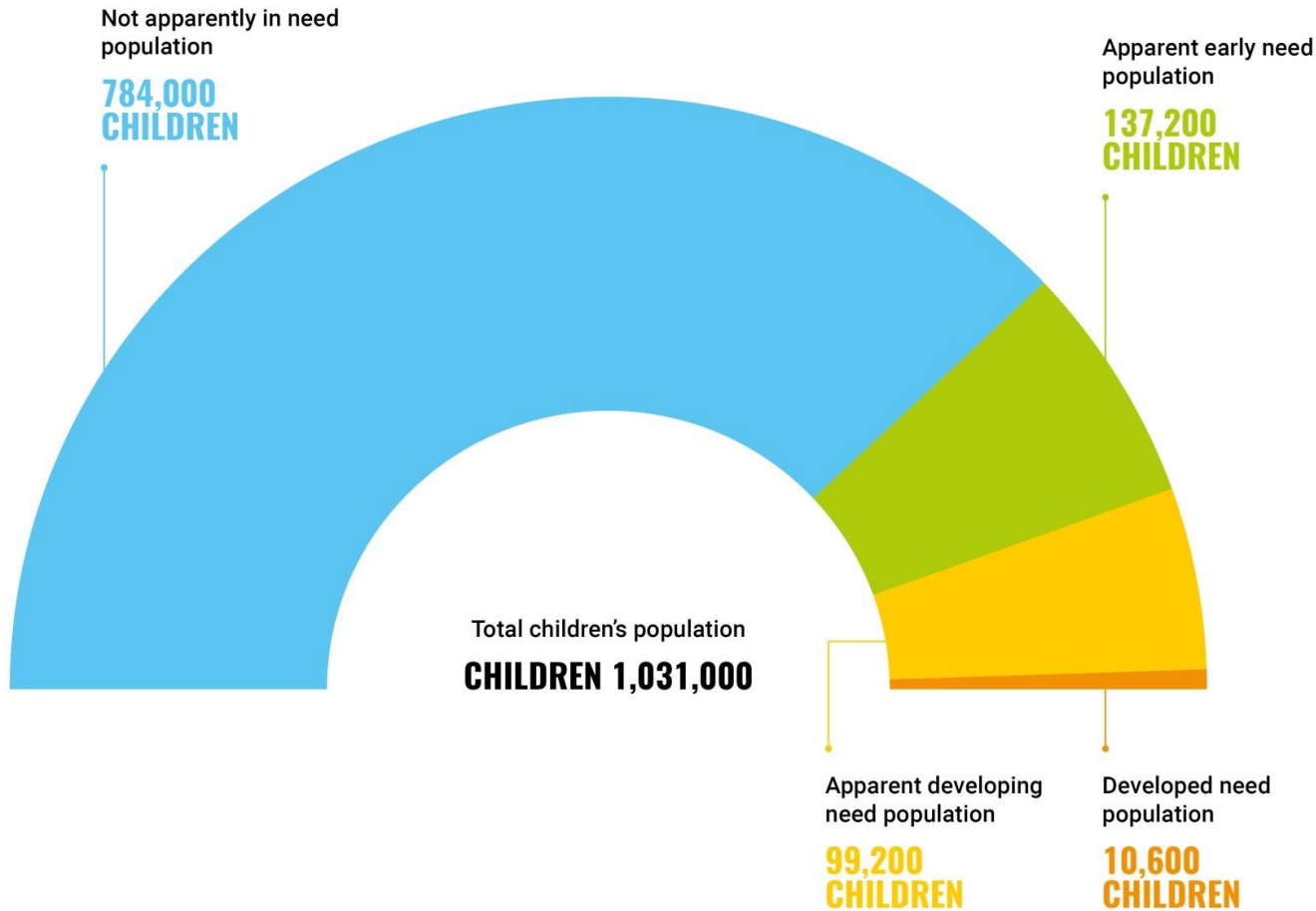
# Exploring the gender pay gap

*"We used multiple robust methods to find similar outcomes, which was strong evidence for policy makers to act."*

Does becoming a mother impact wages for women?

http://women.govt.nz/documents/closing-gender-pay-gap-actions-employers

# Helping children in need

Not apparently in need population
**784,000 CHILDREN**

Apparent early need population
**137,200 CHILDREN**

Total children's population
**CHILDREN 1,031,000**

Apparent developing need population
**99,200 CHILDREN**

Developed need population
**10,600 CHILDREN**

Measuring proxies gives Oranga Tamariki an understanding of each New Zealand child's wellbeing, in the context of them, their family, and their broader community (not in need, early need, developing need, developed need).

- Safety – family violence, OT notifications, injuries…
- Security – social housing register, benefit history…
- Stability – school changes, address changes, multiple caregivers…
- Wellness – family chronic conditions, mental health, hospitalisations…
- Development – truancy, alternative education, NCEA achievements…

# NGOs and Iwi

What can Stats NZ do to better partner with iwi?

# What have we learnt?

# Benefits and limitations

**Benefits:**

- Researchers can tackle previously 'unanswerable' questions
- Longitudinal view
- Cross-sector view
- Geographical views
- Reduced research cost and burden

BUT...

**Limitations:**

- Administrative data quality issues
- High time and skill investment
- Small number studies limitations

# Lessons learnt

- Enabling factors
- "If you build it, they will come"
- Data providers generally willing to make their data available
- Technical skills are a barrier to extracting value
- Early issues with system access and capacity
- Social licence, informed trust
- Flexibility is key

# The next step – IDI 2

# **Problems to solve**

Current and potential IDI users have told us it's difficult to:

- Know **what is being done** already
- Know **what information** is in the IDI(detailed level)
- **Collaborate**/network
- Understand and interact with the **access and use processes**
- Work **iteratively** and at **pace**
- Understand the **quality** of the data
- Get visibility on **interpretation** of data and analysis

# Redesign of IDI production

- A cross government (and wider) working group – focus on strategic direction

- Working directly with users and consumers of IDI to understand the future needs

- Working group takes part in full-day workshop – support team from Stats NZ completes a 3 week sprint in between

# Emerging direction

- Need to ensure a stable and flexible platform to enable growth at the desired rate

- Direction strengthen by focus on 2 key value streams:

  - Better enabling research and insights

  - More focused and deliberate partnering

- Focusing on the experience for the current customer group to enable greater equity of access and ultimately more diverse use
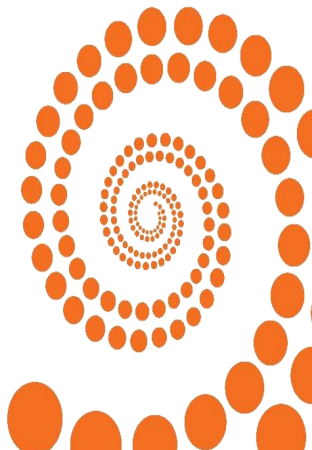
# Questions?

access2microdata@stats.govt.nz

# Additional slides
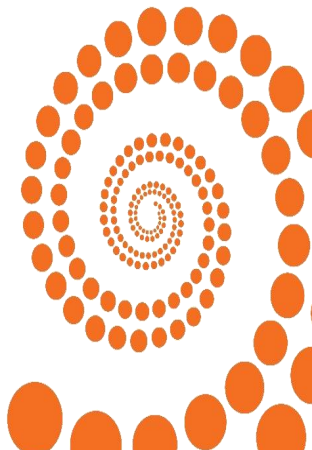
# Technical details

# IDI prototype spine

- The target population for the prototype spine is broadly an 'ever-resident' population
- It includes
  - people born in NZ
  - permanent residents
  - those with visas that allow them to reside, work or study in NZ (including international students and temporary workers)
  - those living and working here without a requirement for a formal visa (e.g. Australians living in NZ)
- It excludes short term visitors (such as tourists)

# Probabilistic linking

- Consider two records and compare the values of the variables in common

- Given the quality of the variables and commonality of the values, assign a probability of these two records belonging to the same person

- The record pair with the highest probability wins