

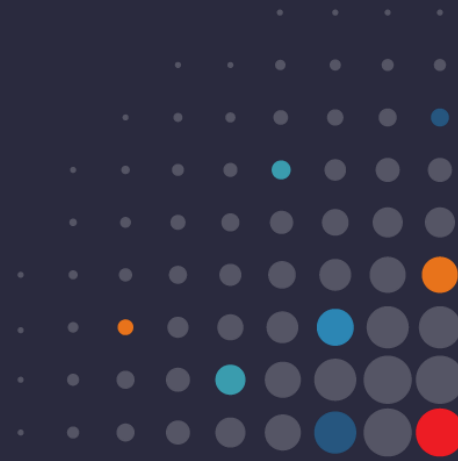
Unlocking the potential of the Integrated Data Infrastructure for research

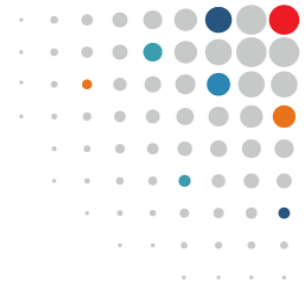
RezBaz June 2026



Social Investment Agency
Toi Hau Tāngata

New Zealand Government
Te Kāwanatanga o Aotearoa





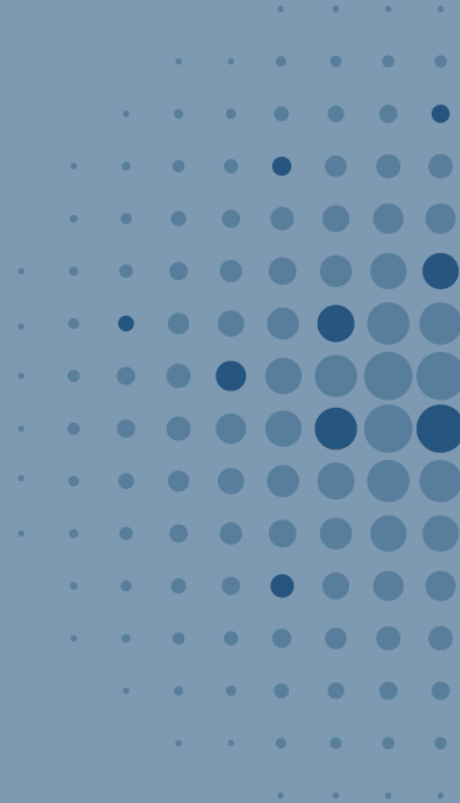
Insight arising from integration

Many insights arise from bringing together information that was once separate

Integrated data is designed for this purpose

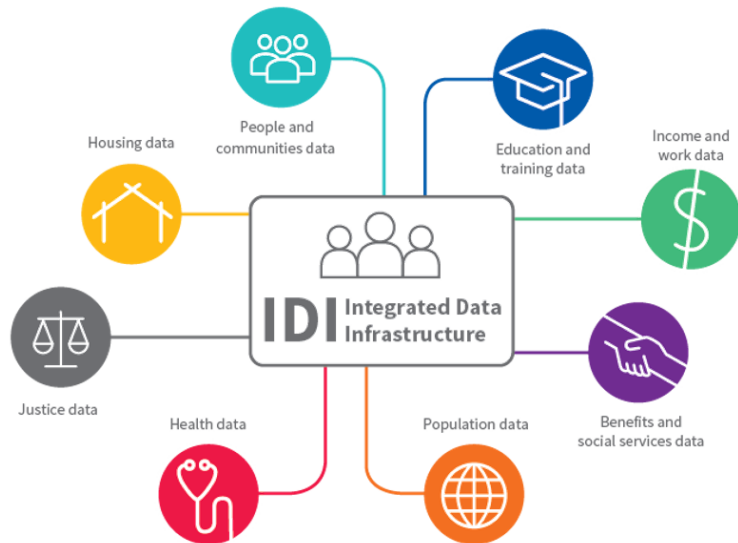
- What is the Integrated Data Infrastructure (IDI)
 - What is available and how it is integrated
 - Data protections in place
- What has been done with it
 - Examples of projects
 - How integration enables questions to be answered
- Some tips, tricks, and resources for using it
 - Significant data wrangling with steep learning curve
 - Range of resources to assist new researchers

What is the integrated data infrastructure?



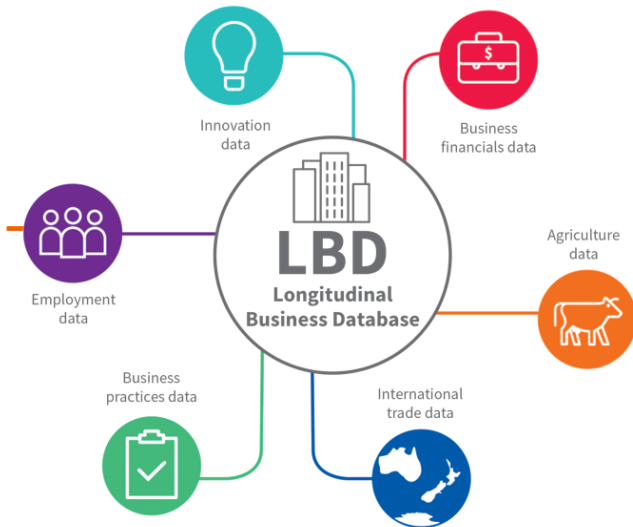
World leading tool for research and analysis

Integrated Data Infrastructure (IDI)

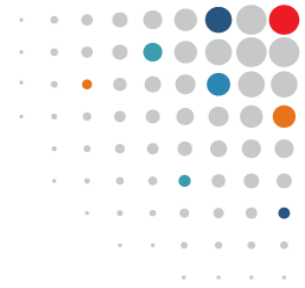


Longitudinal Business Database (LBD)

The IDI and LBD are linked through tax data



<https://www.stats.govt.nz/integrated-data/>



The power of integrated data

Combine information

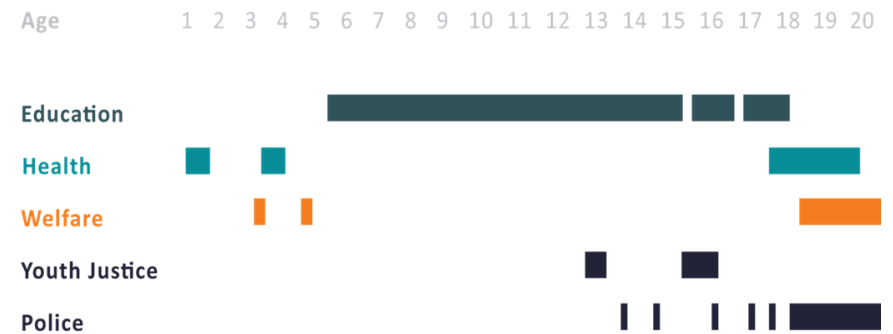
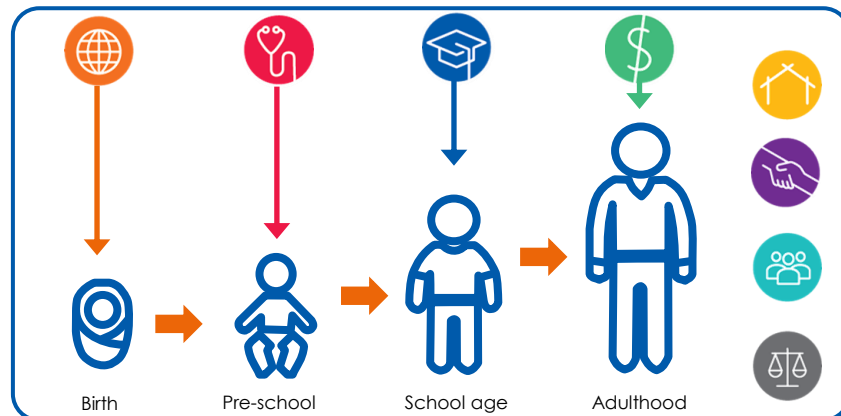
- from multiple domains
- from multiple life stages

Construct detailed cross-sections

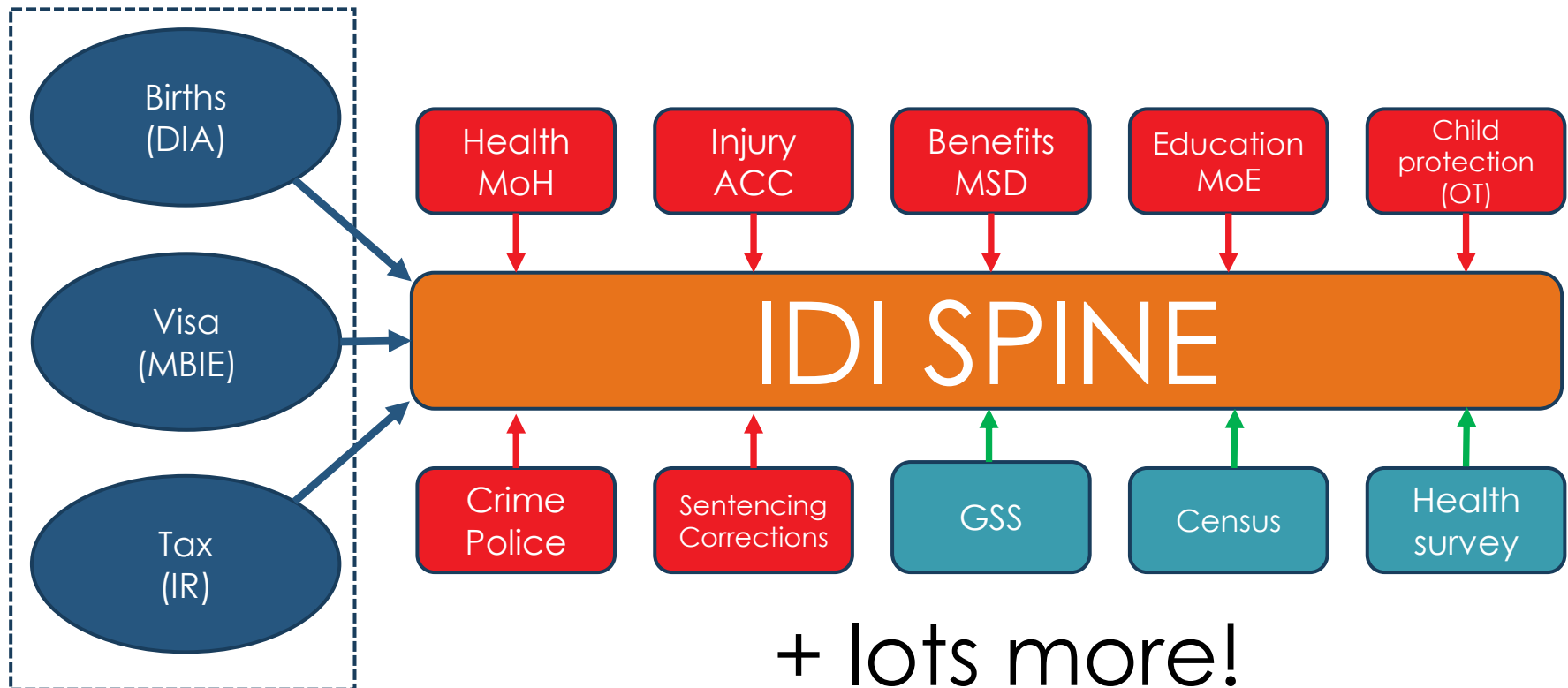
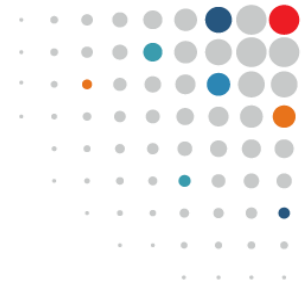
Follow experiences over time

- Interactions between services
- Panel and longitudinal data

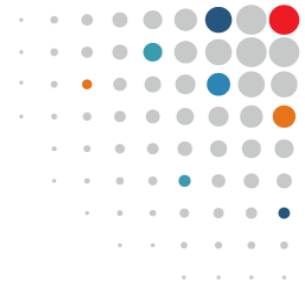
Predict or track outcomes



A list of all identities – the IDI Spine



Deterministic vs Probabilistic linking



Name: Joseph Blogs
DOB: 31 May 1971
Passport Number: ABCDEF
IRD Number: 123456



Health Records



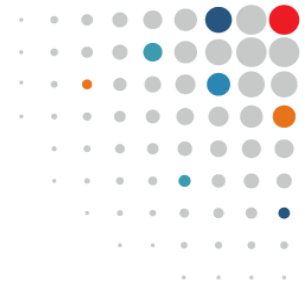
Name: Joe Blogs
DOB: 31 May 1971
IRD Number: 123456



Name: Jo Bloggs
DOB: 31 May 1971



Name: Jo Blogs
DOB: 31 May 1971



'Five Safes' keep integrated data safe

Safe People

Only approved researchers can access or view microdata.

Safe Projects

Data can only be used for research projects in the public interest.

Safe Settings

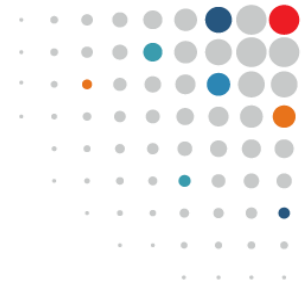
Research takes place in secure data labs which Stats controls.

Safe Data

Access is granted only to the data that is needed for the research.

Safe Output

Confidentiality rules protect against privacy breaches.
All output is checked by Stats NZ to confirm it is safe.



Data is de-identified

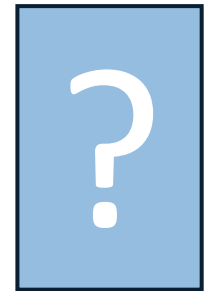
Information supplied to Stats NZ

- **Name:** Star Thinker
- **Date of birth:** 29 February 1933
- **IRD:** 123-123-123
- **NHI:** 0123456789
- **Address:** 123 Enlightenment Terrace, Researchville

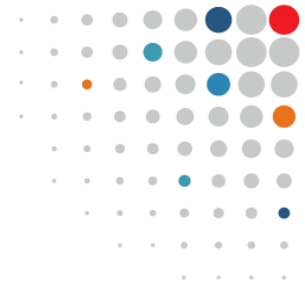


Information visible to researchers

- **snz_uid:** 4545454545
- **Birth month:** May
- **Birth year:** 1981
- **snz_ird_uid:** 111111
- **snz_moh_uid:** 22222
- **address_uid:** 99999
- **Meshblock:** 4507



Confidentiality rules limit data release



Microdata output guide specifies output rules

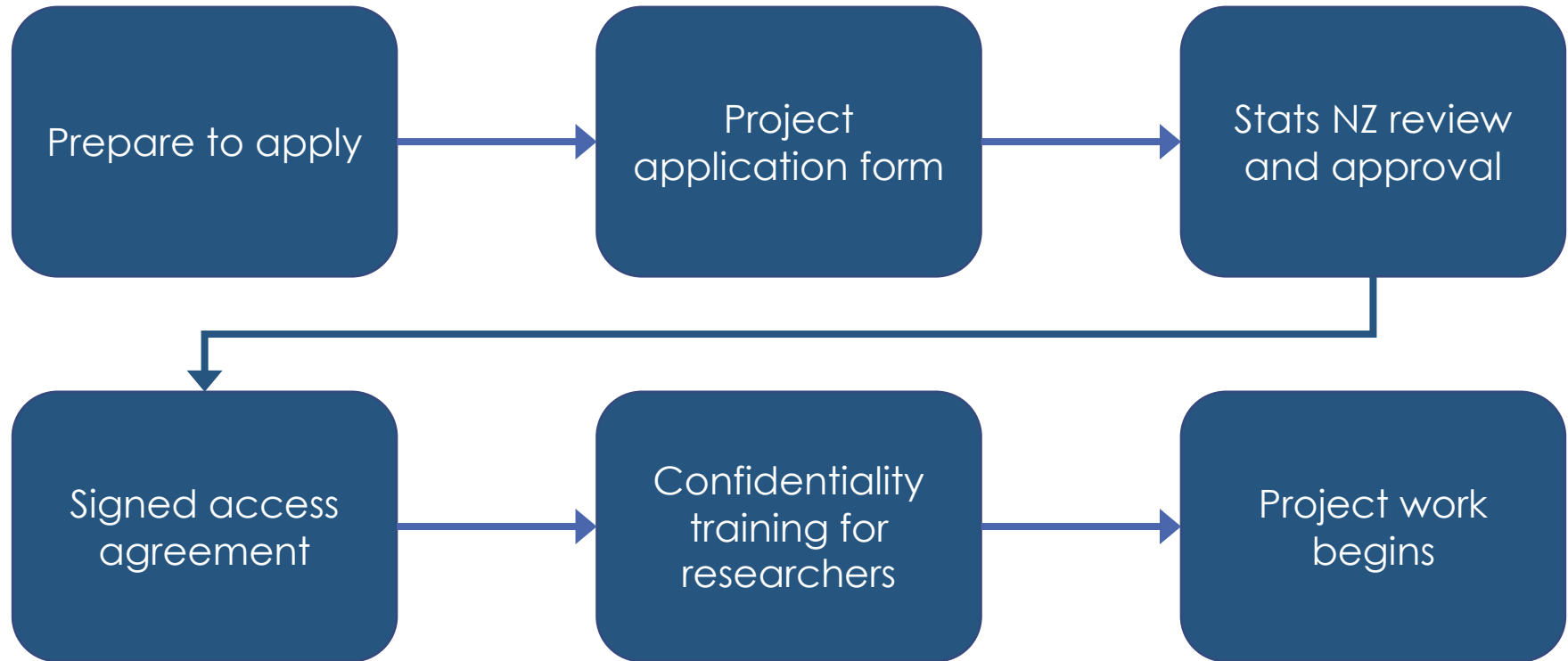
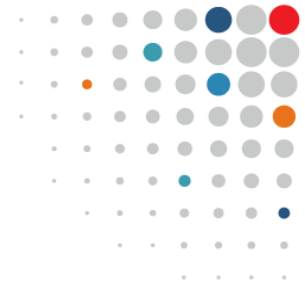
Four most common rules:

- Random rounding of counts to base 3
- Counts of fewer than 6 people are suppressed
- Totals for fewer than 20 people are suppressed
- Values that reflect a single organization are suppressed

Example process

Output	Raw	Released
People in Researchville	62	60
Total income for people in Researchville	\$999999	\$999999
Academics in Researchville	17	18
Total income for academics in Researchville	\$777777	Suppress
Award winners in Researchville	5	Suppress
Total income for award winners in Researchville	\$555555	Suppress
Employees of Top Research Inc	8	Suppress

Applying for access



Apply to use microdata for research

www.stats.govt.nz/integrated-data/apply-to-use-microdata-for-research/

Not perfect – has strengths and limitations



Rich range of data



Skilled and in-demand roles



Secure access



Enforced privacy protections



Range of data quality and documentation shortcomings



Technical capability and knowledge requirements

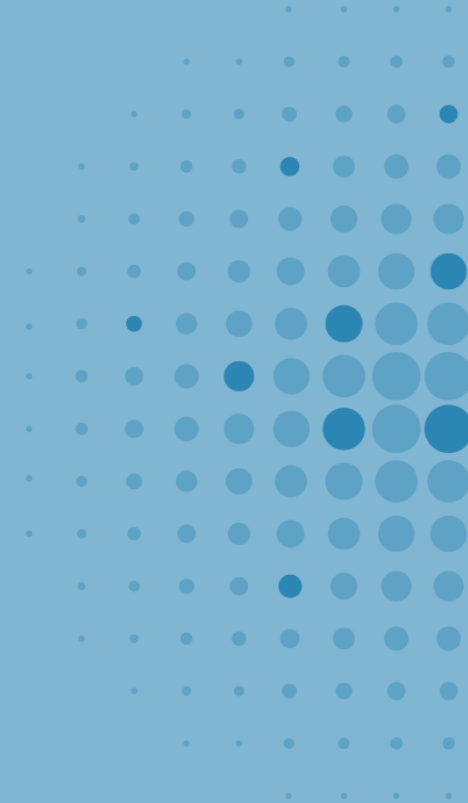


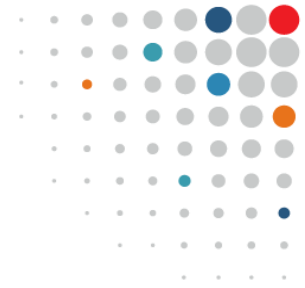
Administrative process
Barriers to collaboration



Small number studies limited

What has been done with integrated data?





Types of questions the IDI is good at

Descriptive

Inferential

Predictive

Overlaps

What don't I know about my clients?

Finding people who have an interaction with agency A (or other characteristic) and looking at what other agencies know about them.

Unmet needs

'Who else should we provide services to?'

The inverse of overlaps – who are the people who we don't see in agency A, but they look like agency A clients?

Impact

'Are we making a difference?'

The richness of data in the IDI can help in identifying comparison groups; the longitudinal nature allows for follow-up.

Lifecourse

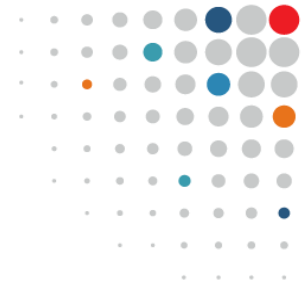
'What leads to the outcomes we want (or don't want)?'

Longitudinal analysis that could be defined by an experience/outcome at the start, middle or end of period.

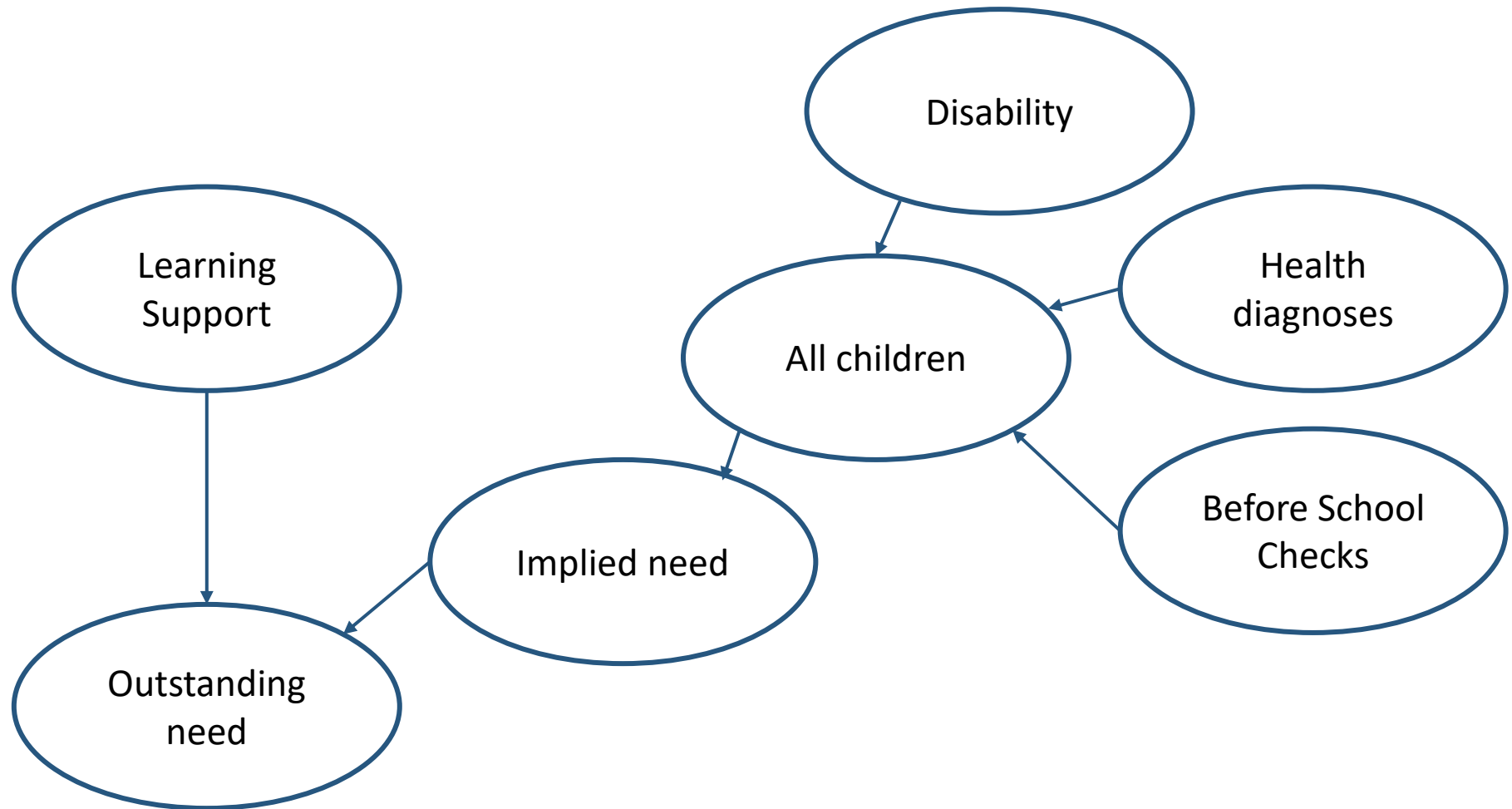
Simulation

'Where should we take action? What would be the likely impact?'

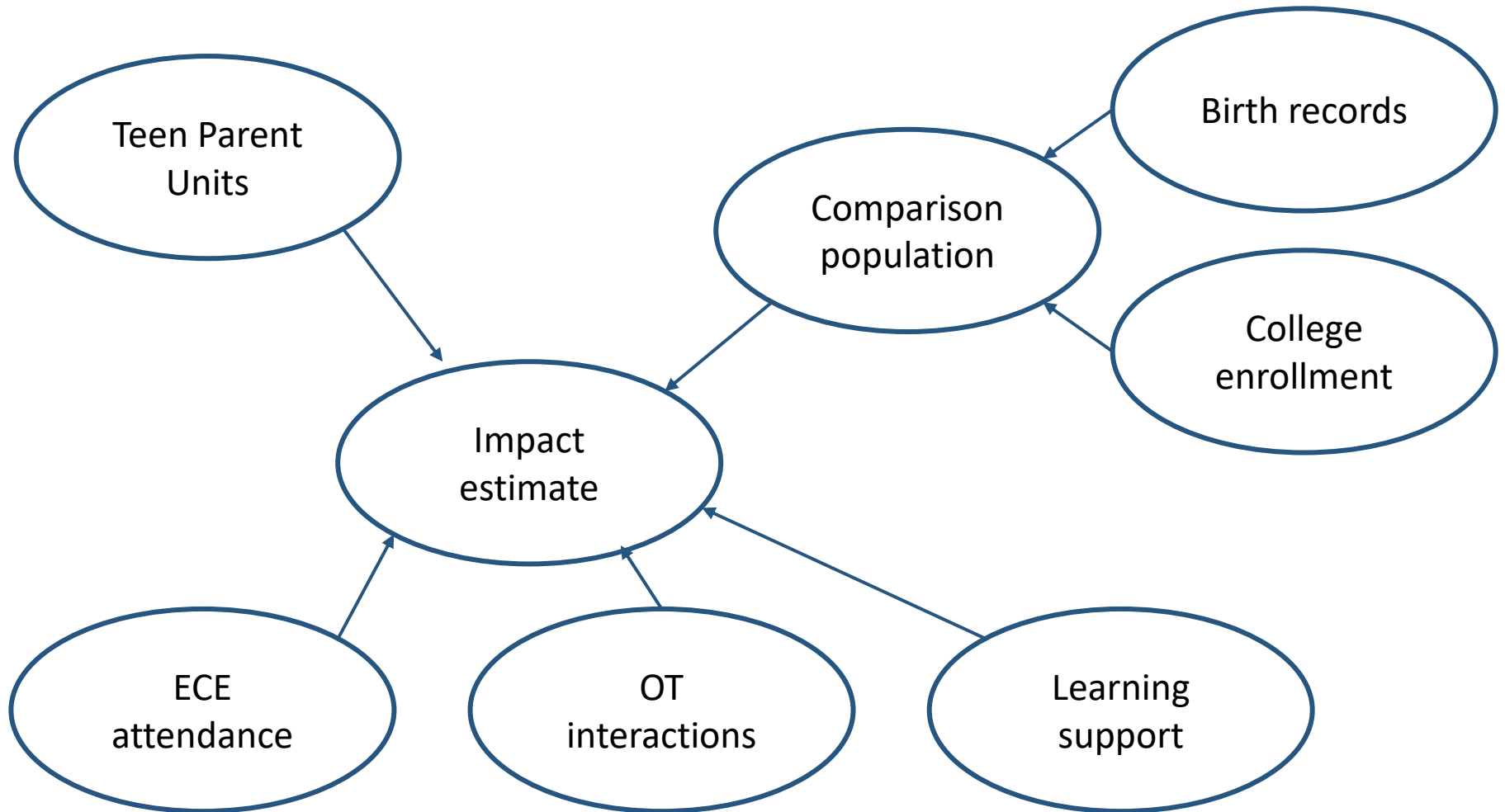
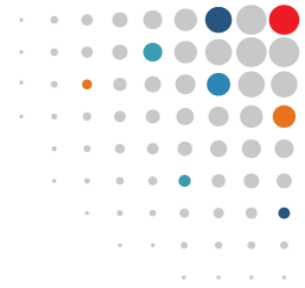
Many models exist in the IDI that can be used for forecasting or 'what if' analysis.



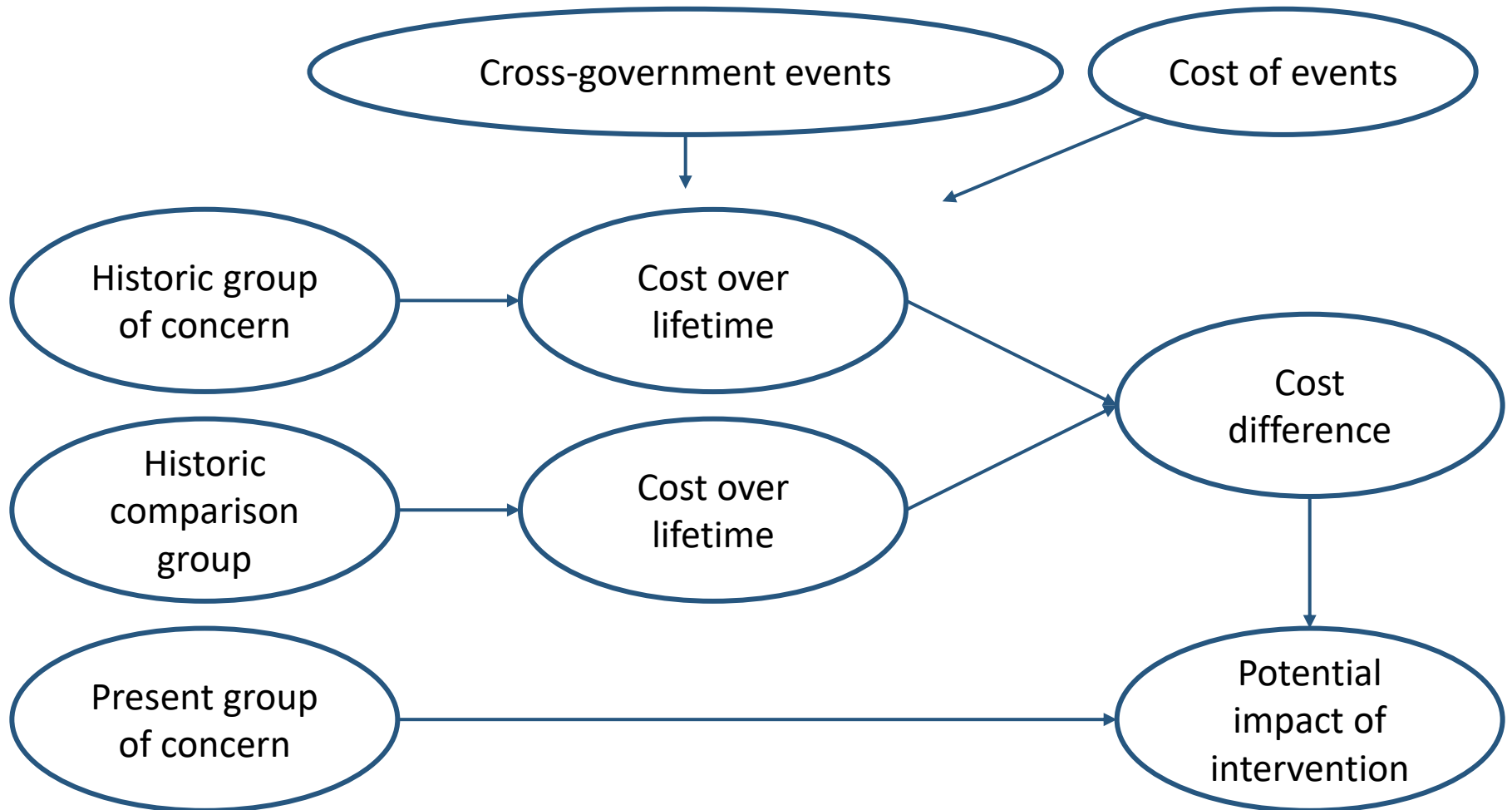
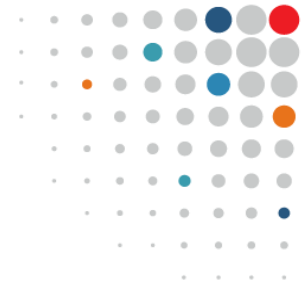
Highest learning needs review



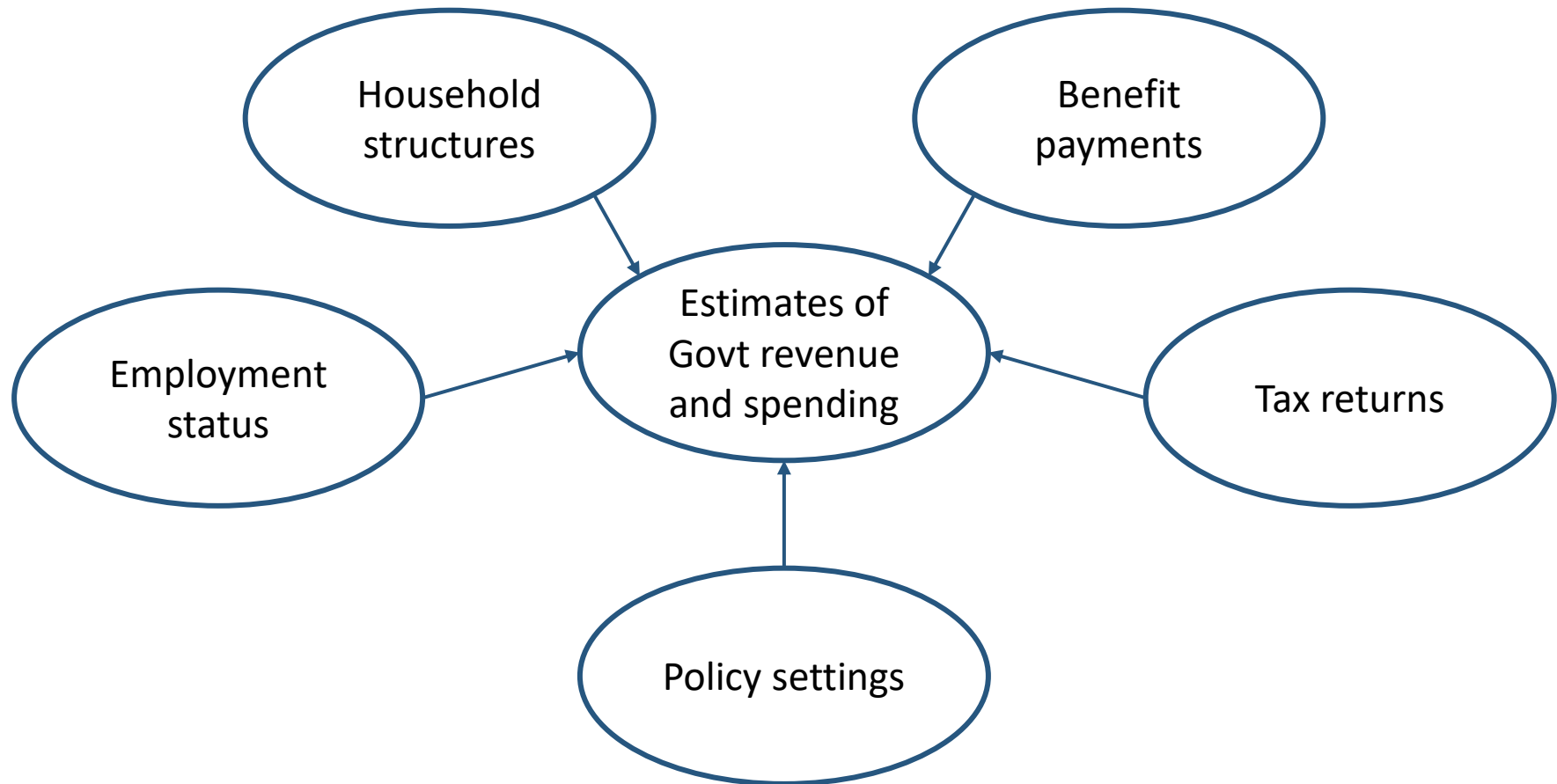
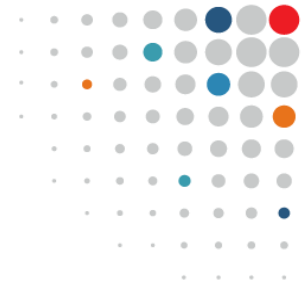
Long-term impact of teen parent units



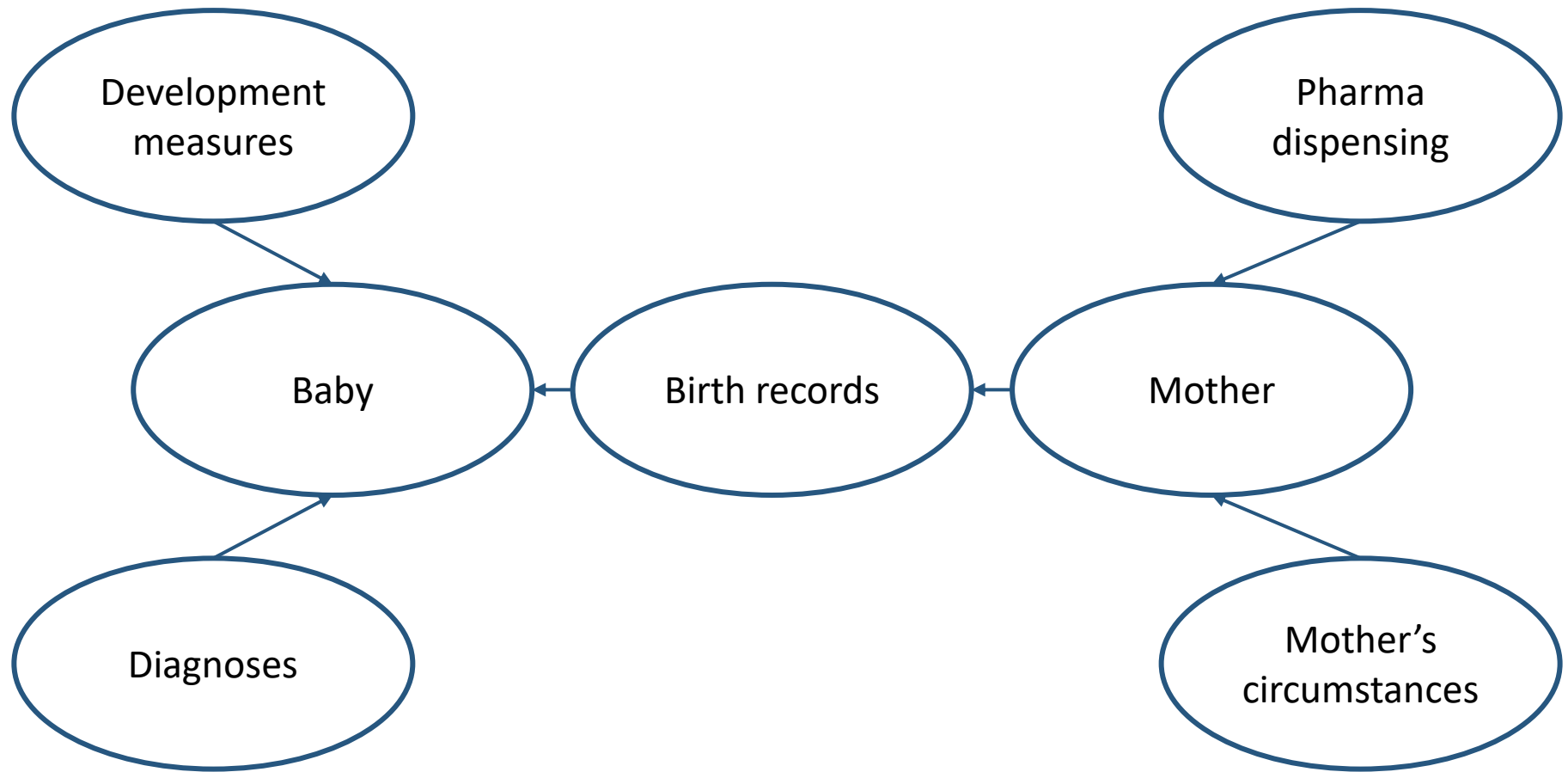
Changes in trajectory



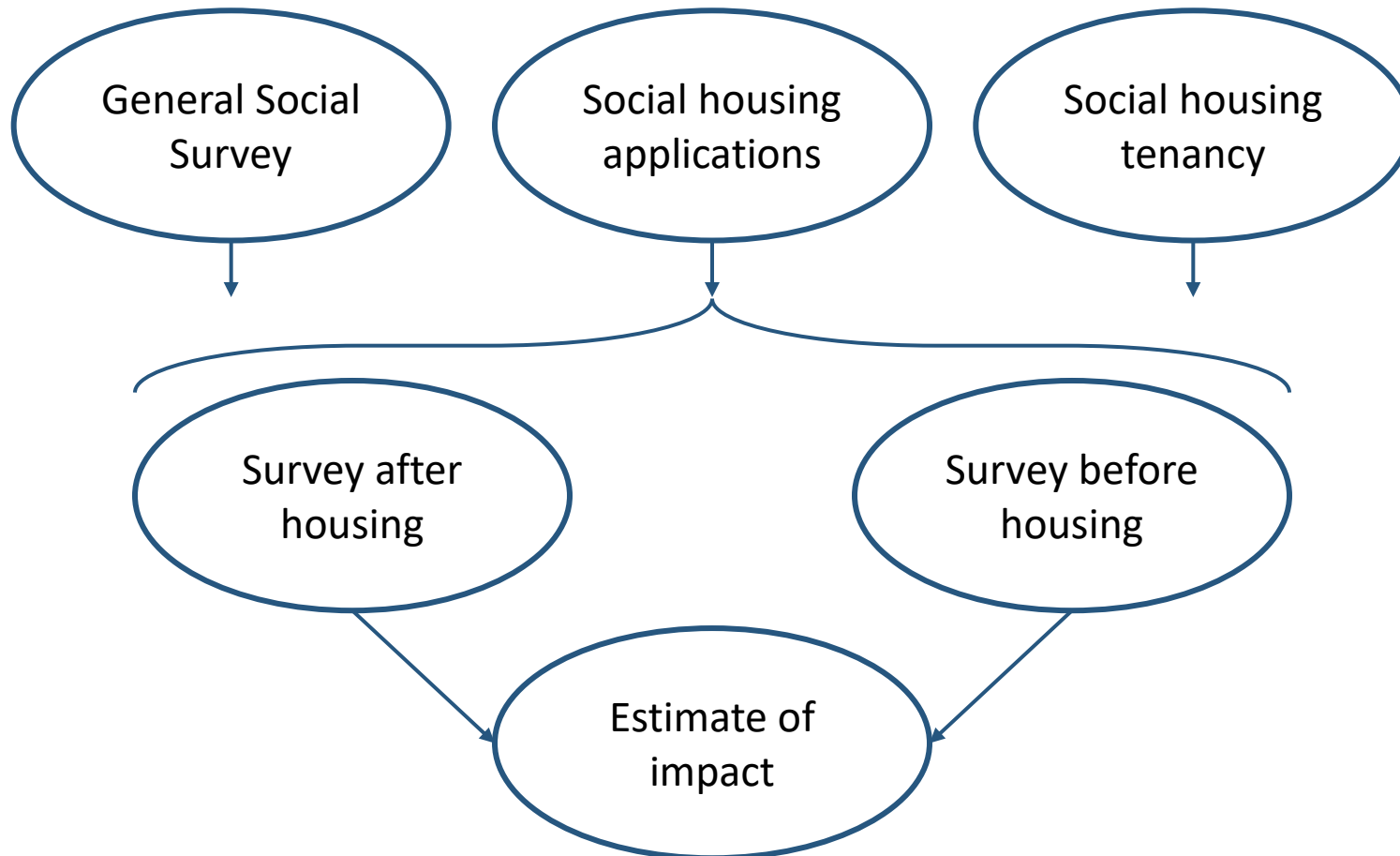
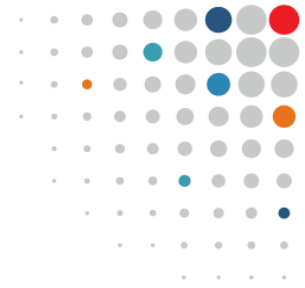
Tax policy microsimulation modelling



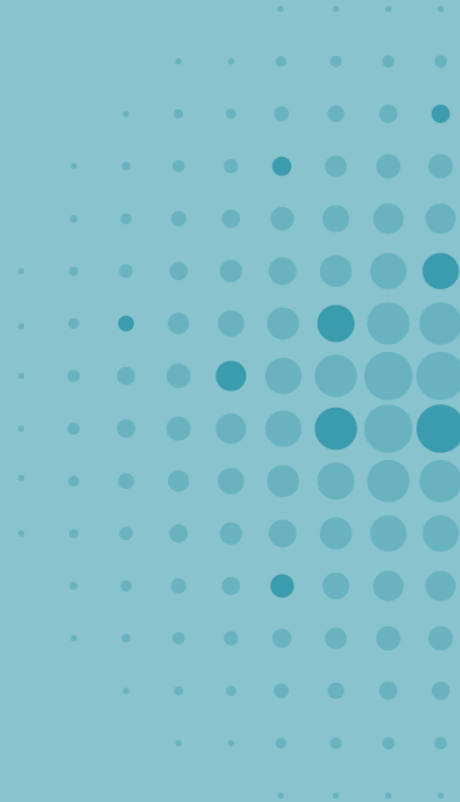
Effect of maternal antibiotic use of babies



Wellbeing impact of social housing



What advice makes it easier to use integrated data?



Lifecycle of an IDI project (our experience)



Determine policy and research question

Define population and period of interest

Create indicators and build dataset

Actual analysis

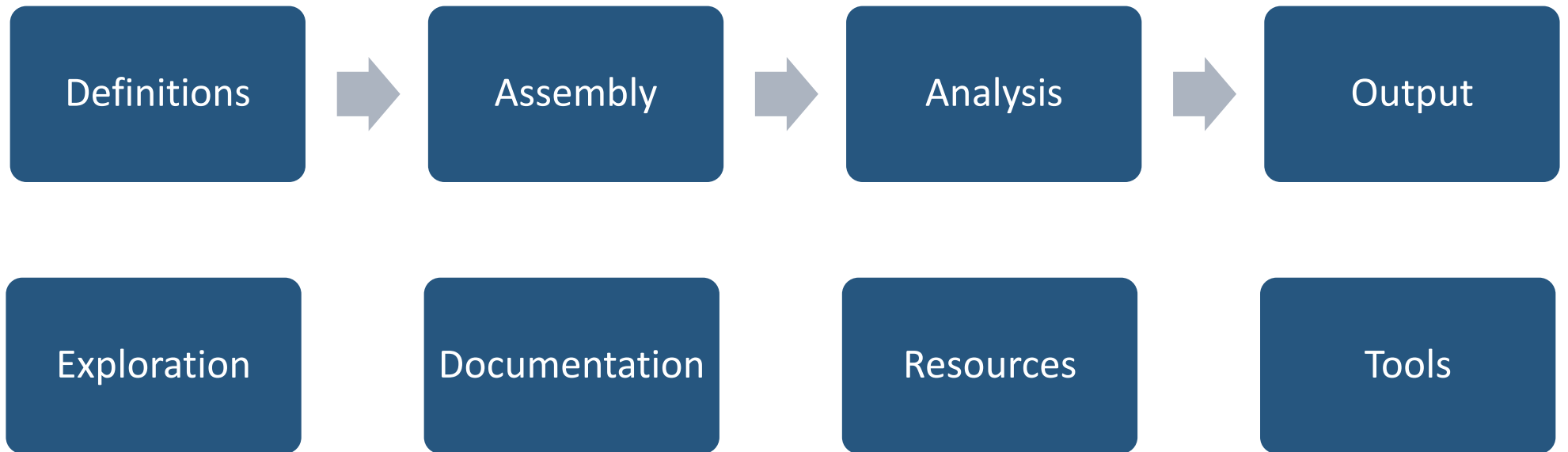
Sense-making and implications for policy

Communication of findings

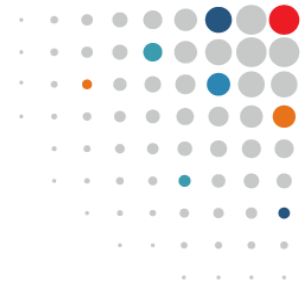
← In IDI →



Structure projects using consistent patterns



Start with a small project



April 2022

IDI exemplar project

Guidance and training



SOCIAL WELLBEING AGENCY | TOI HAU TĀNGATA
TE ROUROU TĀTARITANGA
Information for Social Services and Wellbeing
New Zealand Government

Analysis	agency rename	9 months ago
Assembly	agency rename	9 months ago
Data exploration	agency rename	9 months ago
Definitions	agency rename	9 months ago
Output	agency rename	9 months ago
Resources	Add files via upload	8 months ago
Tools/Dataset Assembly Tool	agency rename	9 months ago
_Checked	agency rename	9 months ago
_For Checking	agency rename	9 months ago
.gitignore	agency rename	9 months ago
LICENSE	Initial commit	9 months ago
README.md	agency rename	9 months ago

README GPL-3.0 license

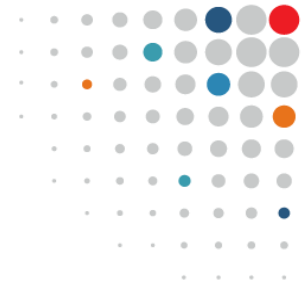
IDI exemplar project

An end-to-end example IDI research project for training and encouraging good practice.

Overview

New Zealand's Integrated Data Infrastructure (IDI) enables incredible research opportunities. However, it can be an intimidating environment to work in for unfamiliar researchers. This exemplar guides new researchers to the IDI through a simple end-to-end project - focused on the practical aspects of managing a project and manipulating the data. The project reflects our current best practice, and we hope that it provides a useful guide for researchers to learn from.

Reference material



Recalibrate expectations

Example task 1:

Compare self-reported life satisfaction against every other measure in the General Social Survey (GSS).

Non-technical perspective:

Concern that large number of crosstabs will be time consuming to create.

Analytic approach:

Quick and straightforward.

Only one input table, already arranged for analysis. Repetitive processing done by computer not researcher.

Example task 2:

Count the number of benefit recipients with children who have diabetes.

Non-technical perspective:

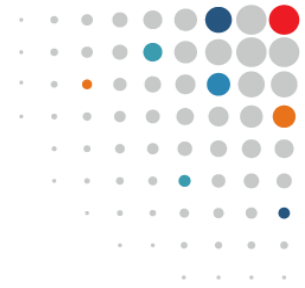
Straightforward as only a single number, benefit receipt, children, and diabetes are all unambiguous concepts.

Analytic approach:

Very challenging.

Diabetes must be constructed from a range of source tables. Multiple ways to define parenting status – may need to test and compare approaches.

Review metadata resources



IDI Search version 1.0.2

Home About Quick Stats Help

Enter search term to filter results Search Clear

Data Supply Agencies (25) CSV JSON

Name
Accident Compensation Corporation
Auckland City Mission

Page 1 of 13

Collections (106) CSV JSON

Name	Agency
ARCOS	University of Auckland
Centre of Innovation and Entrepreneurship Participation	University of Auckland
NZ Rugby Representatives	NZ Rugby

Page 1 of 36

Datasets (1046) CSV JSON

Name	Collection / Agency
Client acc_clean.clients	IDI ACC Injury Claims data Accident Compensation Corporation
Claims acc_clean.claims	IDI ACC Injury Claims data Accident Compensation Corporation
Claims historic acc_clean.claims_historic	IDI ACC Injury Claims data Accident Compensation Corporation
Medical codes acc_clean.medical_codes	IDI ACC Injury Claims data Accident Compensation Corporation
Addresses acc_clean.addresses	IDI ACC Injury Claims data Accident Compensation Corporation

Page 1 of 210

Variables (50049)

Name	Dataset / Collection
1st medical data	IDI LHO REGIONS 2020

Welcome to the IDI Search App

The IDI Search App allows researchers to search for variables that are available in the IDI and, in some cases, metadata about these variables. The app uses data from IDI variables and Data Dictionaries shared with us by Stats NZ. The data are stored in a database which can then be searched using the web app. For help navigating the app, click **Help** in the top right corner.

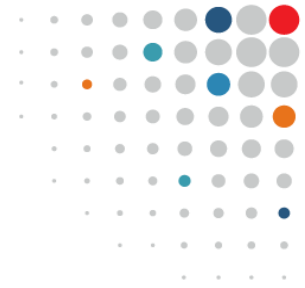
Use the search box to enter terms to filter. To search multiple terms, prefix each word with a plus (+) sign. For example, to search for records that contain both the words "income" and "employment", enter "+income +employment". See Help for more information.

App updated 27 August 2024 [see changelog](#)
Database updated 4 April 2024
Latest refresh April 2024

Proudly supported by

<https://idisearch.terourow.org/>

Video series are available



The image displays a collage of YouTube video player screenshots. The primary focus is on three video series:

- Data Lab Orientation - second video series (1...)**: A playlist of 6 videos with 219 views. The description states: "This video series includes recorded demonstration of how the environment in the data lab looks for researchers...more".
- IDI Demonstration Analysis – series three...**: A playlist of 8 videos with 313 views. The description states: "Stats NZ uses the Five Safes to ensure the protection of IDI data. None of the data used for this series of...more".
- Submitting data lab output**: A video titled "1. Data Output Submissions: Checking Overview" with a duration of 4:25. The video content shows a dark blue background with the Stats NZ logo and the text "Submitting data lab output". A subtitle at the bottom reads: "This video is part of a series on submitting data lab output." The video player interface shows it is 0:10 into the video.

Other visible video thumbnails include:

- S2E1 Remote Desktop**: Stats NZ • 124 views • 1 year ago. Duration: 9:31.
- S2E2 Where to find things in the Data Lab - Stats NZ**: Stats NZ • 93 views • 1 year ago.
- S3E1 IDI Demonstration analysis – Stats NZ**: Stats NZ • 143 views • 1 year ago. Duration: 18:04.
- S3E2 IDI exploration – Stats NZ**: Stats NZ • 146 views • 1 year ago. Duration: 1:26:19.

Build on existing tool and resources

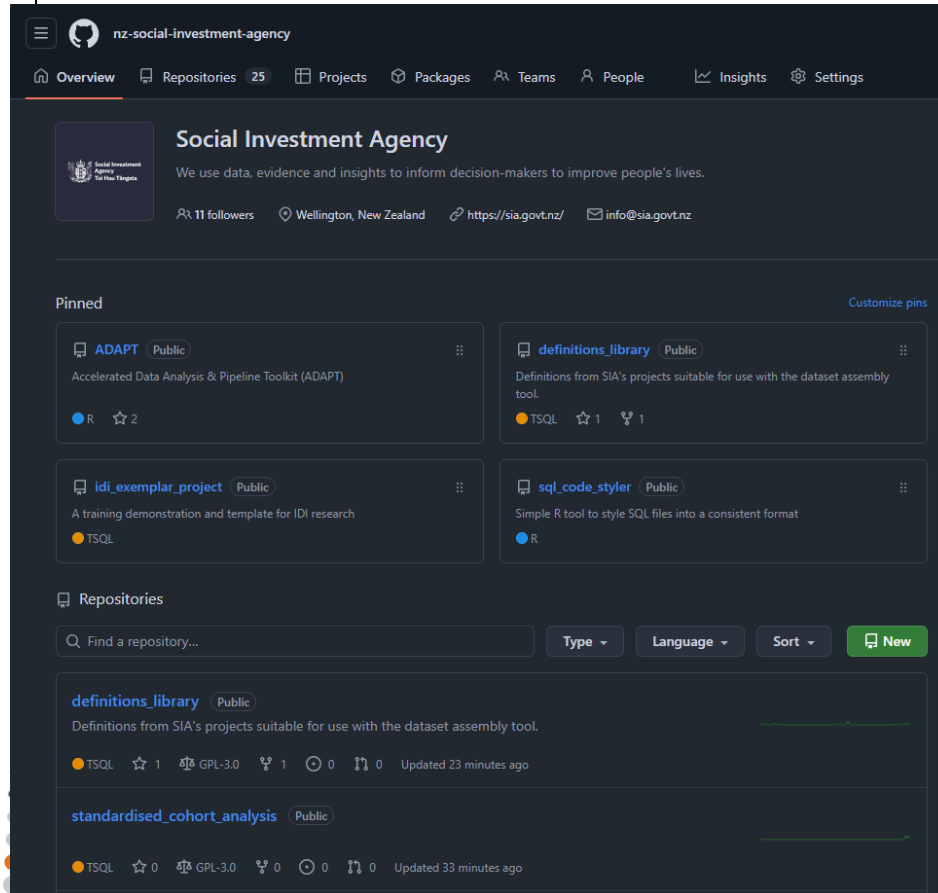


Accelerated Data Analysis & Pipeline Toolkit (ADAPT)

User Guide

March 2026

New Zealand Government
Te Kāwanatanga o Aotearoa

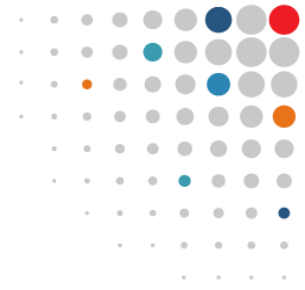


<https://github.com/nz-social-investment-agency>

Connect with the research community

The screenshot displays the Stats NZ website's community forum. The header includes the Stats NZ logo and a search bar with the text "Browse Tags". A left sidebar contains navigation options: "About", "Topics", "Docs", "More", "CATEGORIES" (with sub-items like "Networking and Collab...", "Datasets and Variab...", "Data Lab Guidance", "Skill Share", "Code Modules", and "All categories"), and "TAGS" (with sub-items like "sql" and "All tags"). The main content area features a blue banner with a welcome message and a reminder about privacy. Below this is a navigation bar with "categories", "tags", "Categories" (selected), "Latest", "Bookmarks", "Unanswered", and a "New Topic" button. The forum is organized into several topic cards: "Networking and Collaboration" (Te Mahi Whakawhanaunga me te Pāhekoheko), "Datasets and Variables" (Ngā Kāpuinga Raraunga me ngā Tāupe), "Data Lab Guidance" (He Aratohu Taiwhanga Raraunga), "Skill Share" (Te Tiritiri Pūkenga), "Code Modules" (Ngā Kōwae Uho), "Commons Platform" (Pūhara Tūmatanui), and "News and Announcements" (Ngā Rongokōrero me ngā Pānui). Each card lists specific sub-topics with corresponding color-coded icons.

<https://idcommons.discourse.group/>



Plan for confidentiality rules

Design research within rules

- Estimate population size
- List subgroups you want to analyse
- Will every subgroup be large enough?
- Random rounding adds noise
- Will your results be robust with this noise?
- Track entity counts through analysis, difficult to add retrospectively

Make output process easy

- Stats NZ check 80+ submissions every week
- A small amount of extra effort on each output submission adds up to days of extra effort
- Spend a little more time ensuring submission is correct and clear
- Save the checker time and save yourself delays
- Watch the video series on good output practice before your first submission

Distinguish between different table layouts

Tidy rectangular source

ID	region	age	income
1	north	younger	200
2	north	older	400
3	north	younger	100
4	south	older	200
5	south	younger	100
6	south	older	0
7	south	older	400
8	south	younger	300
9	south	younger	400
10	south	younger	400
11	north	older	100
12	north	older	300
13	north	older	0
14	north	younger	400
15	north	younger	200
16	north	older	300

Long-thin results

region	age	count	total income
north	older	5	1100
north	younger	4	900
south	older	3	600
south	younger	4	1200
-	older	8	1700
-	younger	8	2100
north	-	9	2000
south	-	7	1800
-	-	16	3800

Presentation results

count	younger	older
north	4	5
south	4	3
total income		
north	900	1100
south	1200	600

**Save early,
save often**

**Your workspace
resets every Sunday
night (but sometimes
at random intervals!)**



Simon Anastasiadis

info@sia.govt.nz

sia.govt.nz



Social Investment Agency
Toi Hau Tāngata

New Zealand Government
Te Kāwanatanga o Aotearoa

