Office for National Statistics

IDEAS - Integrated Data Enabling Analysis and Statistics

GSS Conference, 22 November 2017

Becky Tinsley

IDEAS Context

Produce coherent and reliable statistical information within less time and using less resources

Faster response to requests for critical information

Better Statistics, Better Decisions

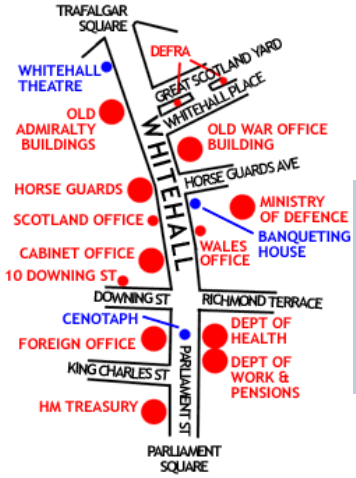
Strategy for UK statistics, 2015 to 2020

Ambitious strategy for UK Statistics

IDEAS Context

Digital Economy Act 2017 – A legislative framework for data access to produce statistical information

World class digital infrastructure

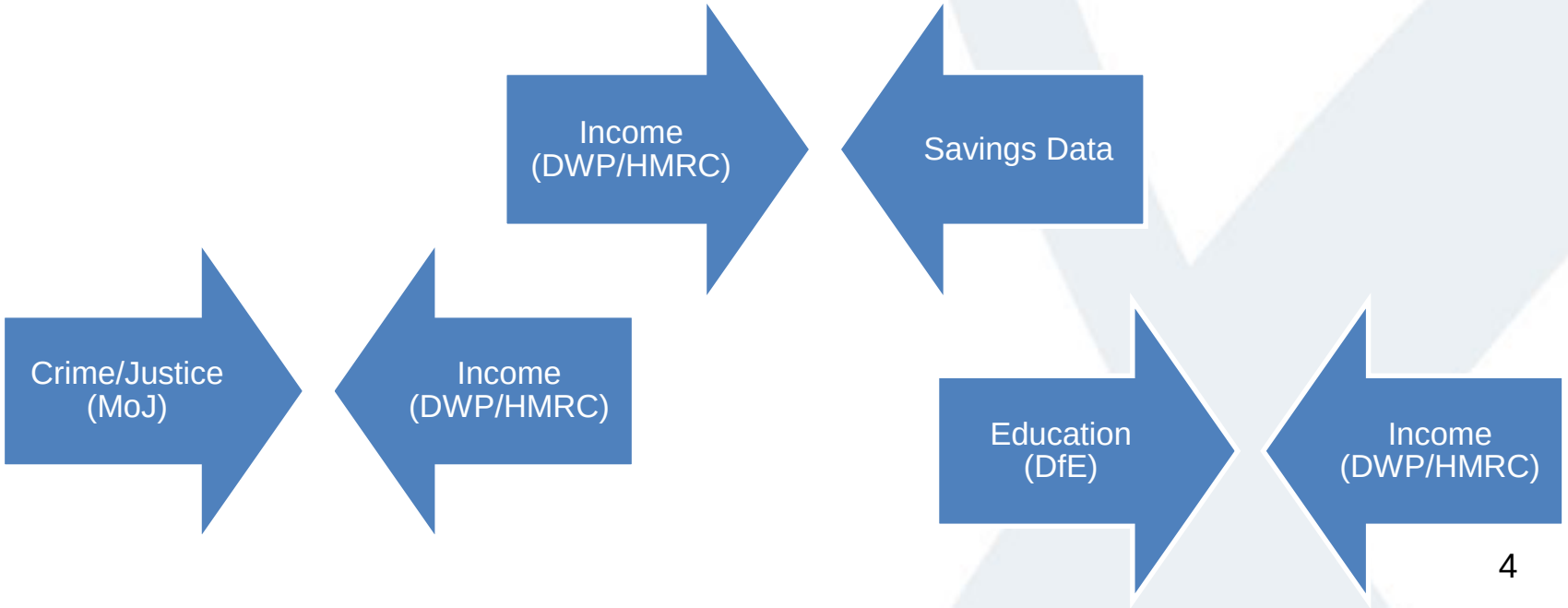


Joined up Whitehall collaboration on shared interests (common measurement of households, income)

Data Linkage

- Joining records from two sources containing different information for the same entity

- Linkage work already carried out across government and providing insight:



Systematising Linkage – Data Integration

What do we mean by data integration?

Holding de-identified datasets in such a way that can be consistently and coherently joined together...

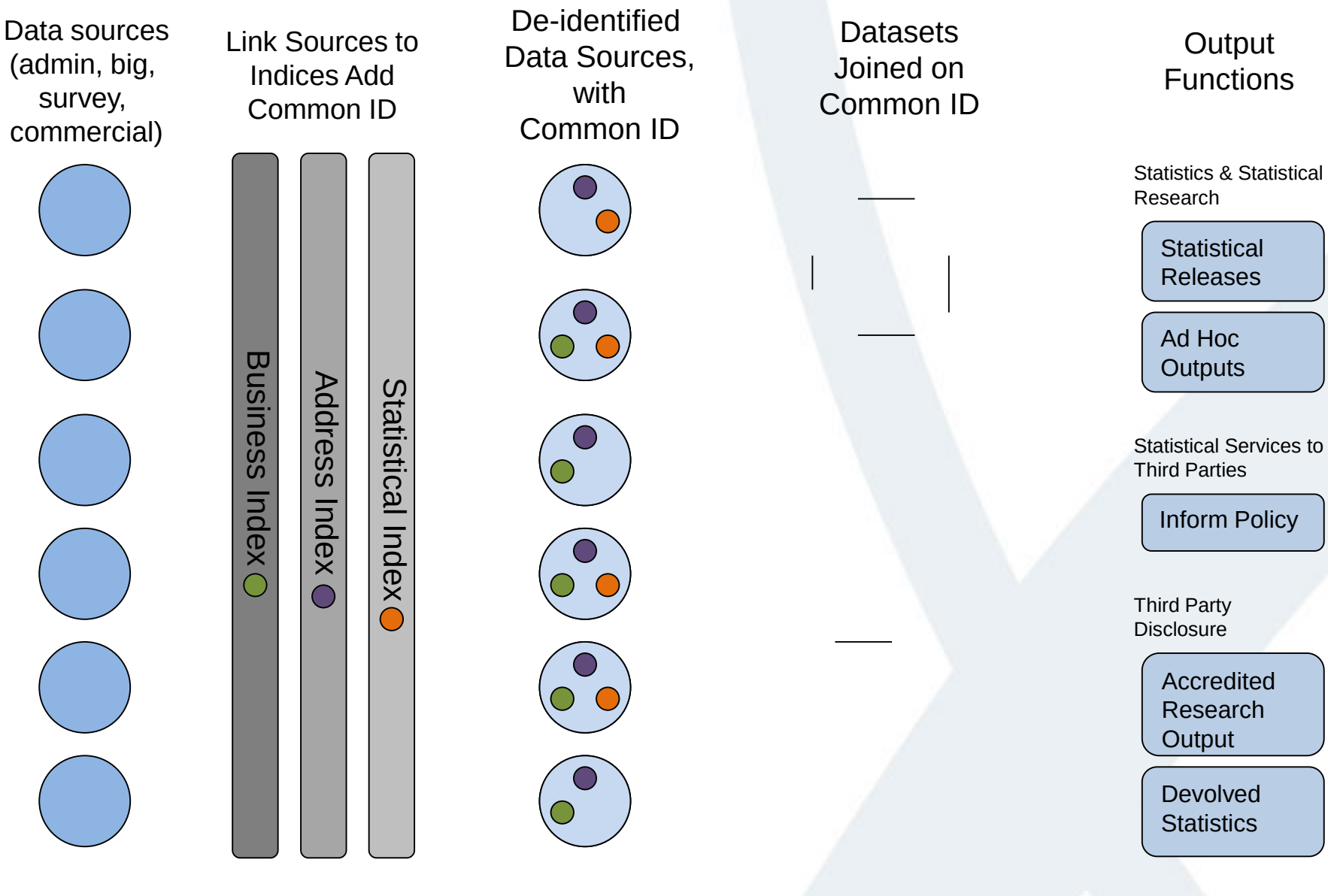
…for a range of different applications

…in a far more timely way than the existing data model…

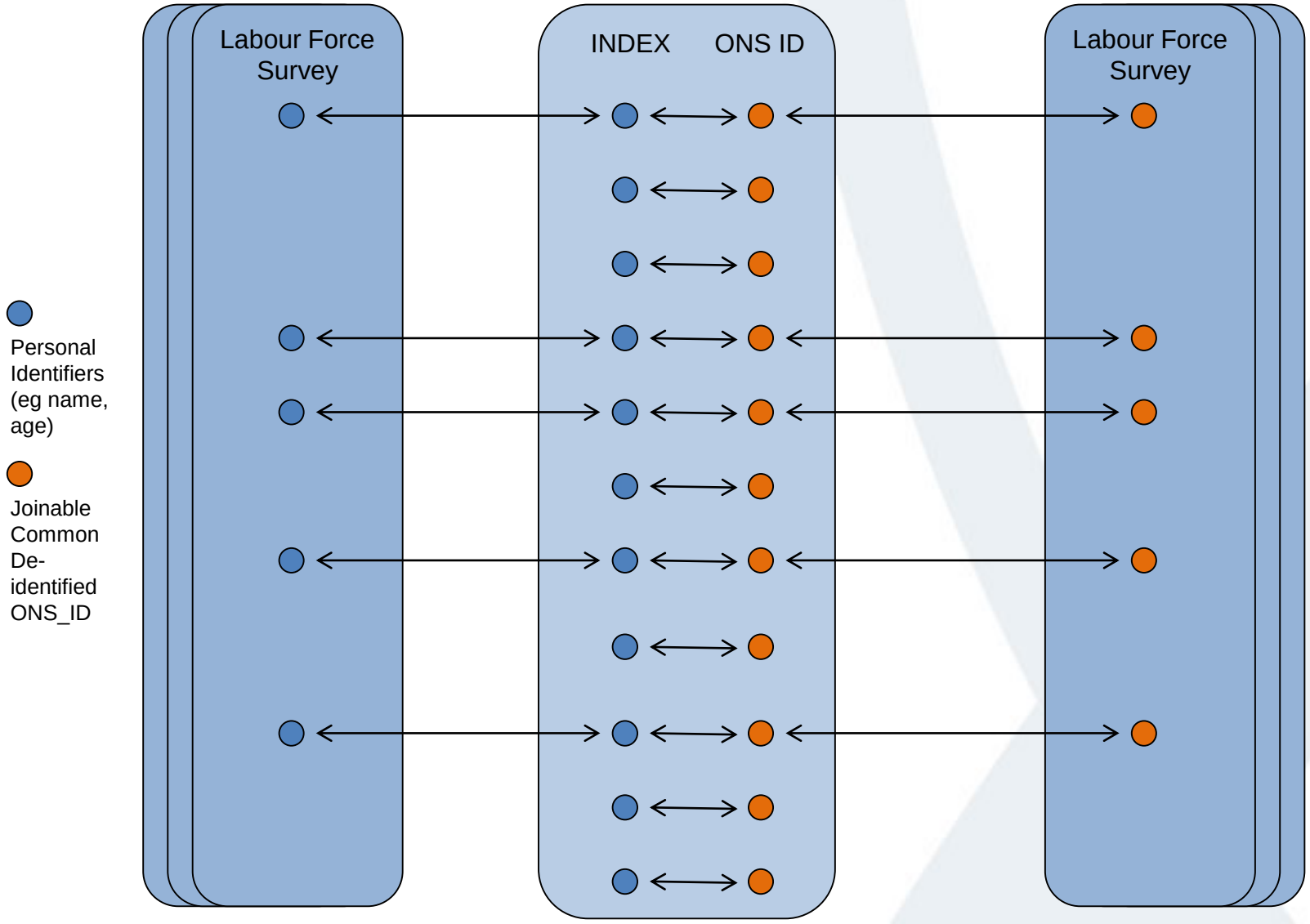
…without being a large dataset of everything

Data Integration – 3 way index

IDEAS - A New Data Model



Creating and Using an Index

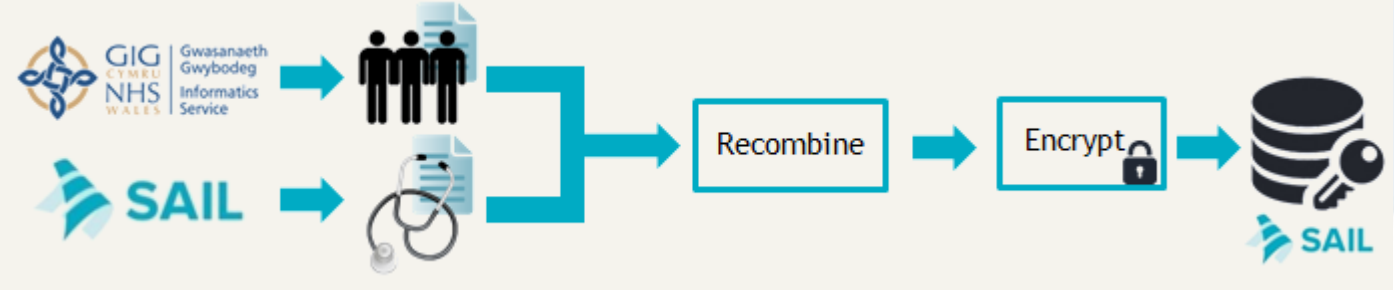


Using IDEAS - Principles

- Only a data model
  - Any application would need to be both legal and ethical
- De-identified at point of use
  - Researchers only ever work with de-identified data
- Separation of duties
  - Maintenance of and linking to index is separated from any working with characteristic information
- Separation of storage
  - No characteristic information stored with index

Similar models

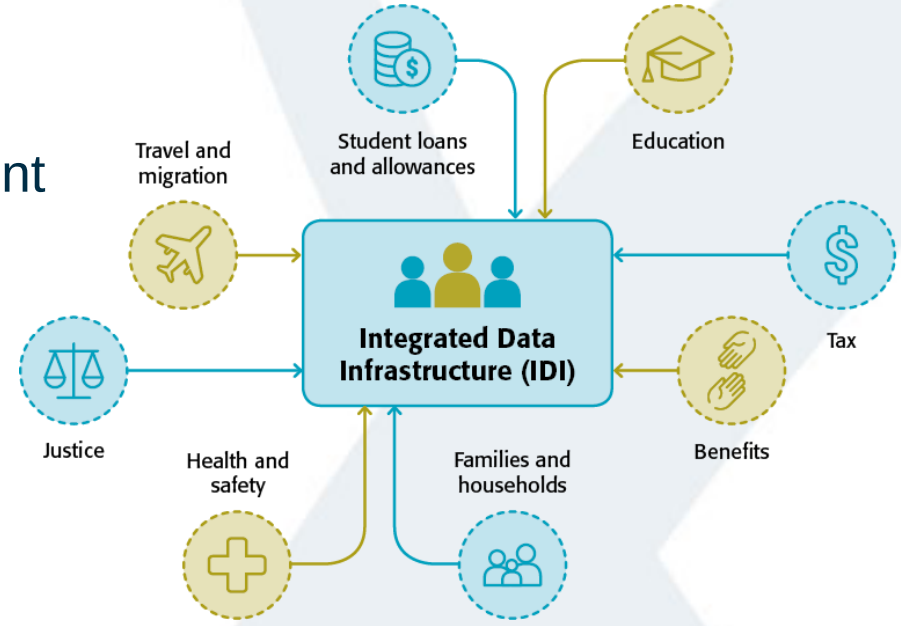
- Established in 2006
- A system which links de-identified data at individual and household level across many health and health related datasets in Wales.
- Offers a secure environment and training to link datasets to answer health related questions
- A research ethics committee retains the ethical oversight of the process and a member of the National Research Ethics Service sits at an independent review panel.



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4139270/

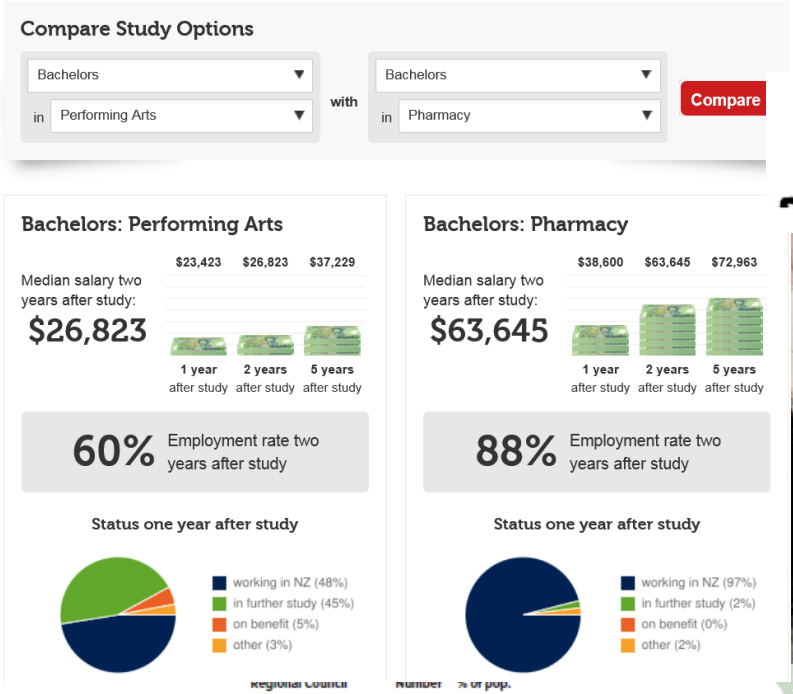
Similar models

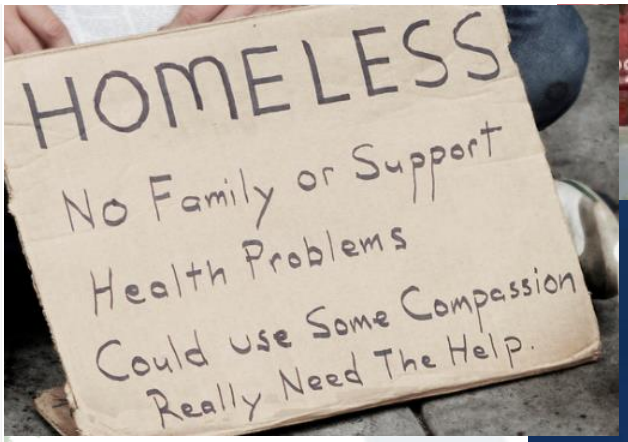
- Established in 2011
- The IDI is a large SQL database holding linked longitudinal microdata about people and households.
- Index creating by linking information about the same identity across multiple data sources in different ways to produce statistics for the public good.
- Operate within the 5-safes framework
- Continuous public engagement leading to informed trust
- Health projects require ethics approval
- Training provided



How the IDI is being used in NZ

Stats NZ
Tatauranga Aotearoa

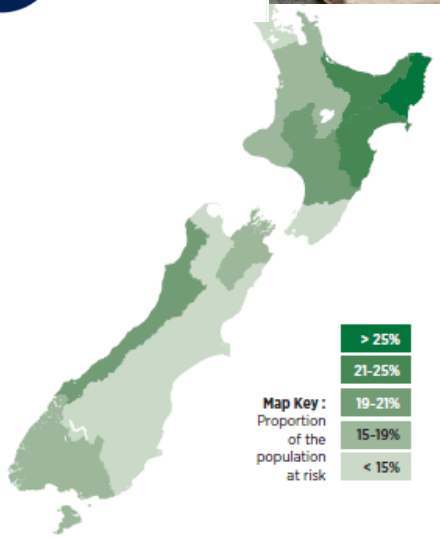




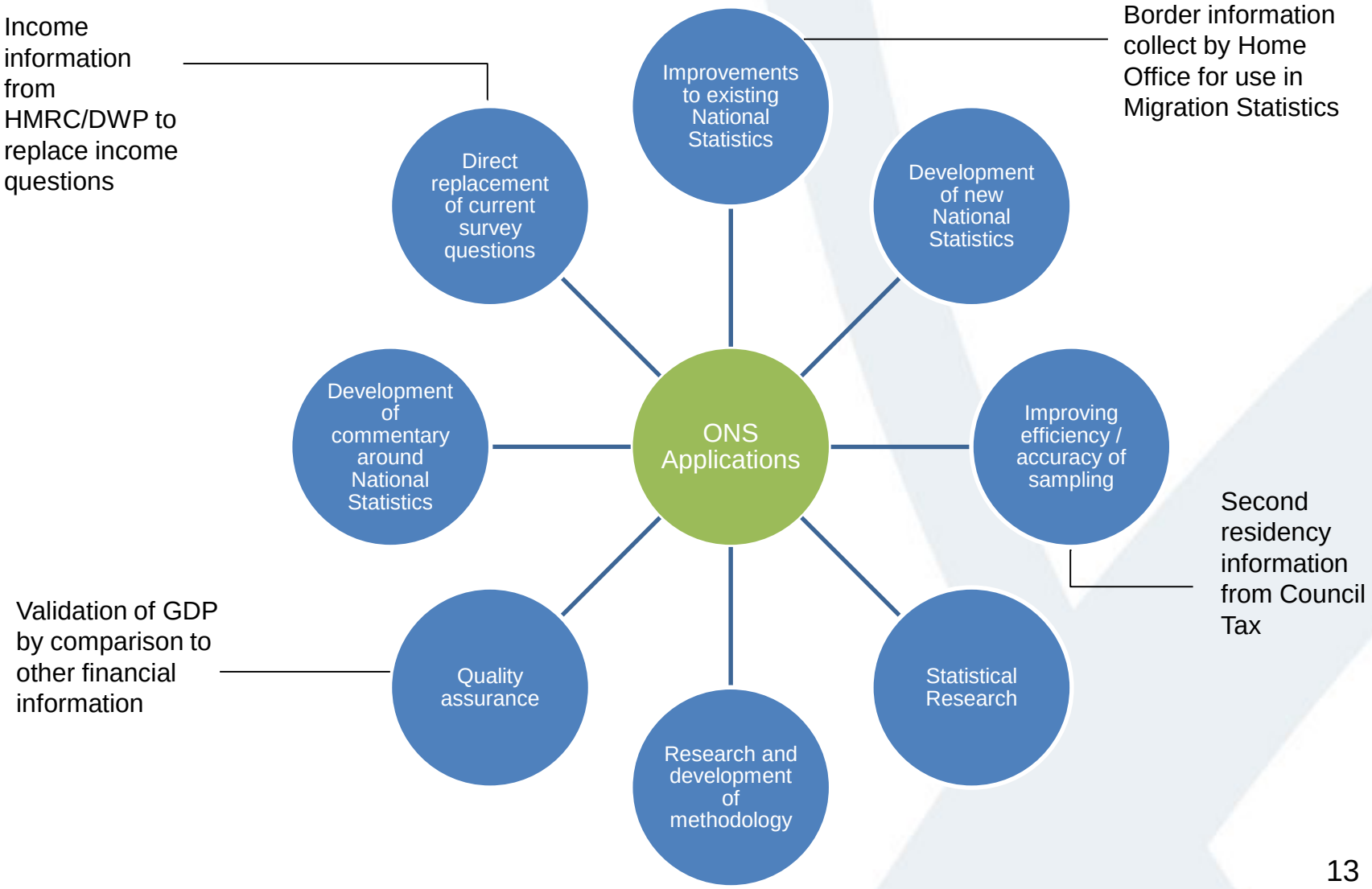
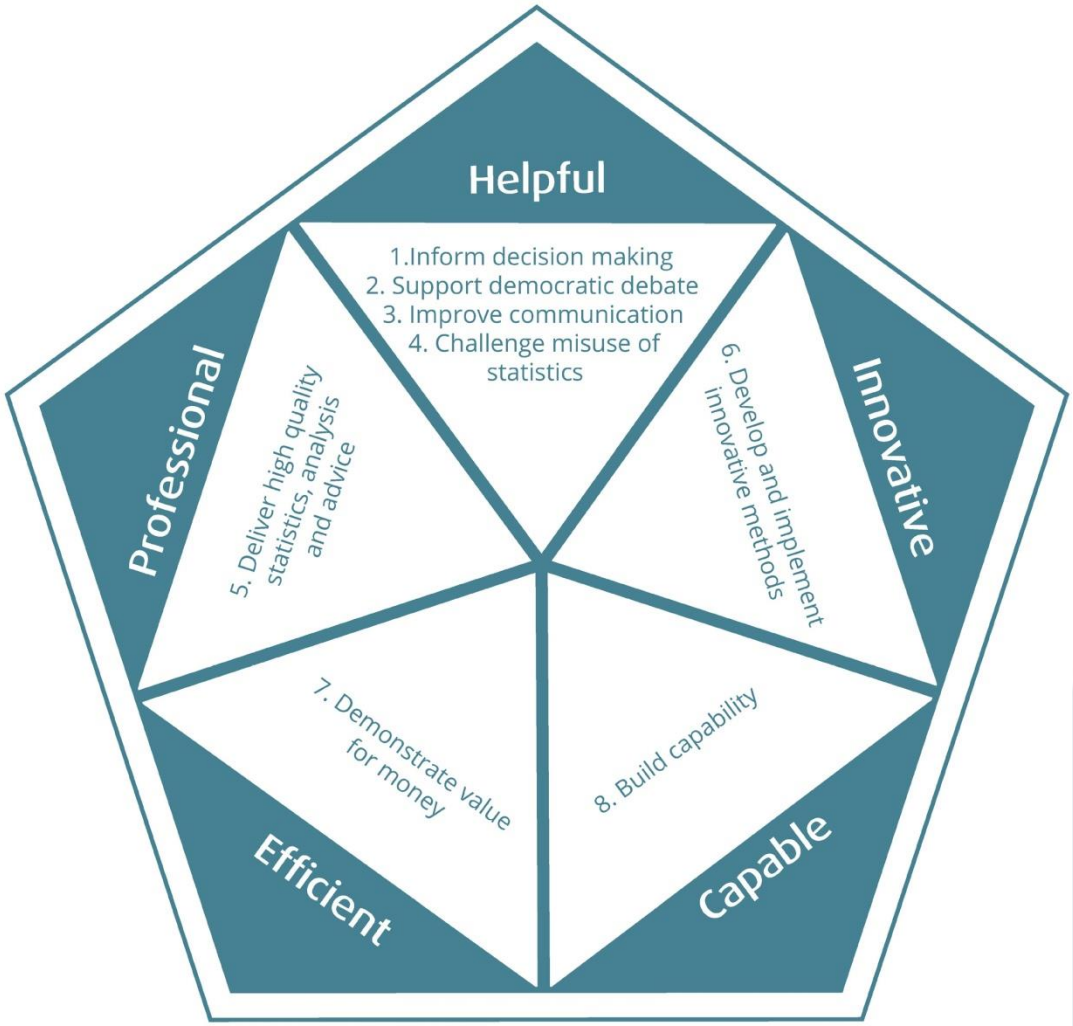
15% MOST AT RISK BY REGION

We know where these young people are:

Regional Council	Number	% of pop.
Gisborne/East Coast	1,944	28.7%
Northland	5,577	27.5%
Hawke's Bay	4,533	22.1%
Bay of Plenty	7,941	21.6%
Manawatu-Wanganui	6,363	20.1%
West Coast	732	19.3%
Waikato	10,107	18.6%
Southland	2,274	18.2%
Nelson	1,086	18.2%
Marlborough	858	17.9%
Taranaki	2,517	17.3%
Canterbury	8,820	12.4%
Wellington	7,932	12.3%
Tasman	642	11.9%
Auckland	22,998	11.5%
(- South Auckland	10,017	15.4%)
Otago	2,940	10.7%
NEW ZEALAND	87,250	15.0%



Integration to benefit - ONS Applications



How does IDEAS fit with the GSS Strategy?