

New Zealand’s Integrated Data Infrastructure

Sarah Minson (Statistics New Zealand)

sarah.minson@stats.govt.nz

Abstract

There is an increasing focus on Statistics NZ as an enabler for research, allowing insights into complex issues and opportunities that cross agency boundaries. To generate these insights, decision-makers require complex joined up data.

Statistics NZ has been undertaking data integration for over 10 years. The early integration projects were for specific purposes, often linking two or three datasets, and each project was kept in a separate environment. In 2011, Statistics NZ created the Integrated Data Infrastructure (IDI) prototype – consolidating previously separate data. At this point Statistics NZ moved from one-off data integration to providing a data integration service.

In 2013, the New Zealand Cabinet agreed to expand the use and reach of the IDI by allowing government agencies to share data – for this a cross-agency data-sharing solution was required.

The Integrated Data Infrastructure (IDI) is a large research database containing microdata about people and households. Data is from a range of government agencies, Statistics NZ surveys including the 2013 Census, and non-government organisations. The IDI holds over 166 billion facts, taking up 1.22 terabytes of space – and is continually growing. Researchers use the IDI to answer complex questions to improve outcomes for New Zealanders.

The presentation will cover:

- What is the Integrated Data Infrastructure (IDI) and how did New Zealand achieve this?
- How we keep the data safe
- How researchers are using the IDI