

## D. Validating official statistics

125. There have been cases where other sources are seen as comparable to official statistics, and when they differ, the official statistics have been challenged. [One example from the United Kingdom](#) shows how the distribution of businesses listed in the "Yellow Pages" telephone directories was compared to the coverage of the statistical business register. A further example concerns comparisons of inflation figures from MIT's Billion Prices Project against official price indices. These examples show that "other sources" are reaching a level of credibility that challenges the role of official statistical organisations.

126. It is possible to use external data sources to determine accuracy of survey results or use survey results to challenge results from alternative data sources of providers of statistics either at the micro level, i.e., linking information from multiple data sources on an individual person or business firm (unit) or at a macro level, i.e., linking information from multiple data sources on a group of people or business firms (units).

127. The issues involved in integrating alternate data sources into the validation processes used for producing official statistics include:

- assessing origin and quality of the source, including trustworthiness and commercial or other interests of the parties exploiting them
- designing processes and modelling techniques which are sustainable and formalised (as ad hoc adjustments to the statistics would be difficult to defend) and
- educating users on proper use and interpretation of information (both the general public and more specific user groups).

128. There are a number of related initiatives:

- [ESS.VIP ADMIN Project](#). This project aims to find ways to optimise the use and accessibility of administrative data sources in the production of official statistics while guaranteeing the quality and comparability of these statistics.
- [ESSnet project on Data Integration](#). This completed project focused on the methodologies and methodological issues of micro data integration.
- [ESSnet project Integration of Survey and Administrative Data](#). The project aimed at developing the knowledge and expertise of participating statistical organisations in the use of integrated survey and administrative Data in the production of official statistics.
- [ESSnet project on macro-integration](#). This project discusses various methods of integrating data sources at aggregated or macro level.

129. The following are paragraphs outline the opportunities that data integration has provided at Statistics New Zealand to validate statistics.

130. The advancement of data integration skills has also led to the creation of [Statistics NZ's Integrated Data Infrastructure \(IDI\)](#). The IDI brings together linked datasets from a range of government agencies (including Statistics NZ's own data collections). The IDI is a large research database containing microdata about people and households and is continually growing. The IDI has paved the way to answer complex research questions to improve outcomes for New Zealanders.

131. Administrative data have been linked to examine and decide on their specific use in the production of official statistics. Inland Revenue data, specifically longitudinal payroll data from the Employer Monthly Schedule (EMS) returns was linked to produce new statistics - filled jobs, worker flows, and total earnings - that measure labour market dynamics at various levels - including industry, region, territorial authority, business size, sector, sex, and age. These statistics provide an insight into the operation of New Zealand's labour market.

132. Data integration has also been used for the improvement of a survey process as illustrated in the linking of the March 2013 Household Labour Force Survey (HLFS) to the 2013 Census data to analyse non-respondents to the HLFS. The project led to the deletion of a non-response adjustment step in the weighting procedure for the HLFS simplifying the HLFS estimation process.

133. Some validation projects involving the use of various administrative data sources have led to recommendations of using these data sources for either benchmarking income survey results, imputation or validation of income statistics rather than using the administrative data sources to replace various sources of income. The administrative data sources need not be integrated to the income surveys when using them for benchmarking or validating income statistics. In cases where data integration will be required for the above immediate uses, a new process - data integration - will need to be designed in the production process.

134. Linking the Census to administrative data sources in the IDI has been instrumental in the realisation of some of the goals of Statistics NZ's [Census Transformation Programme](#). The programme is investigating alternative ways of running New Zealand's future census including the feasibility of using linked administrative data to replace census questions.

135. Data integration has also paved the way in the development of new methods, e.g., new models. One good example is the production of population estimates using administrative data. Bryant and Graham (2015) use Bayesian modelling to estimate, specifically, regional populations in New Zealand based on administrative data on birth and death registrations, tax and NZ international passenger movements.

136. Integrating multiple data sources face a number of challenges. A number of these are described in the following paragraphs.

137. Timeliness of external data sources, unless receipt of data is common and regular, will always be a challenge for linked datasets and for statistics produced from these datasets. These include:

- The promptness in picking up birthed units to an administrative data source. A high number of birthed units not picked up quickly enough by administrative data leads to potential bias in official statistics.
- Timing issues around getting the linking accomplished in time for production.

138. The cooperation of the dataset owner is also another challenge to address. The statistical organisation needs to ensure the continuity and consistency of the quality of the data to be provided. However, contingency plans need to be in place in case the data source becomes unavailable. The statistical organisation may also need to elicit assistance in determining the definition of concepts, classifications or populations in case these need to be redefined to better suit their needs.

139. After quality assessment of an external data source has been undertaken, the next challenge to address is the extent an external data source will be used to meet the statistical need. Are new methods required to convert the external data source into a form useful in the production of a statistical output?

140. Although administrative data may be freely available to a statistical organisation, other external data sources may not necessarily be available for free. Costs may also be a challenge in accessing external data sources. Costs are also incurred in the quality assessments of external data sources and all these costs need to be determined and assessed before proceeding with any data integration project.

141. Another challenge is the resistance to changing any part of a production process that will involve the integration of an external data source especially when current approaches are widely accepted, and well-grounded expertise has been established.

142. The need for standardised processes which are responsive to administrative changes in the data supplied and to new administrative data available to Statistics NZ should also be addressed when using external data to validate official statistics.