Australian Bureau of Statistics

The Rise of Non-Survey Administrative Data for Statistics and Evidence-Building

Dr David Gruen AO
Australian Statistician

13 January 2025

Overview

- ABS is using administrative data to support the production of official statistics to:
  - produce new official statistics;
  - replace existing surveys;
  - replace select data items in existing surveys; and
  - assist with sample selection and survey response follow-up.

Produce new official statistics

- The monthly household spending indicator is based on transactions data from major banks and scanner data from supermarkets via data supply agreements. Covers about 68% of household consumption, compared to 30% for the Retail Trade survey. Will replace the Retail Trade survey in mid-2025.
- The monthly business turnover indicator is based on turnover data supplied by firms that remit Business Activity Statements once a month to the Australian Taxation Office (ATO – equivalent to the IRS).
- Payroll jobs from Single Touch Payroll (STP). The ATO receives payroll information from employers through STP-enabled accounting software each time the employer runs its payroll.

Replace existing surveys

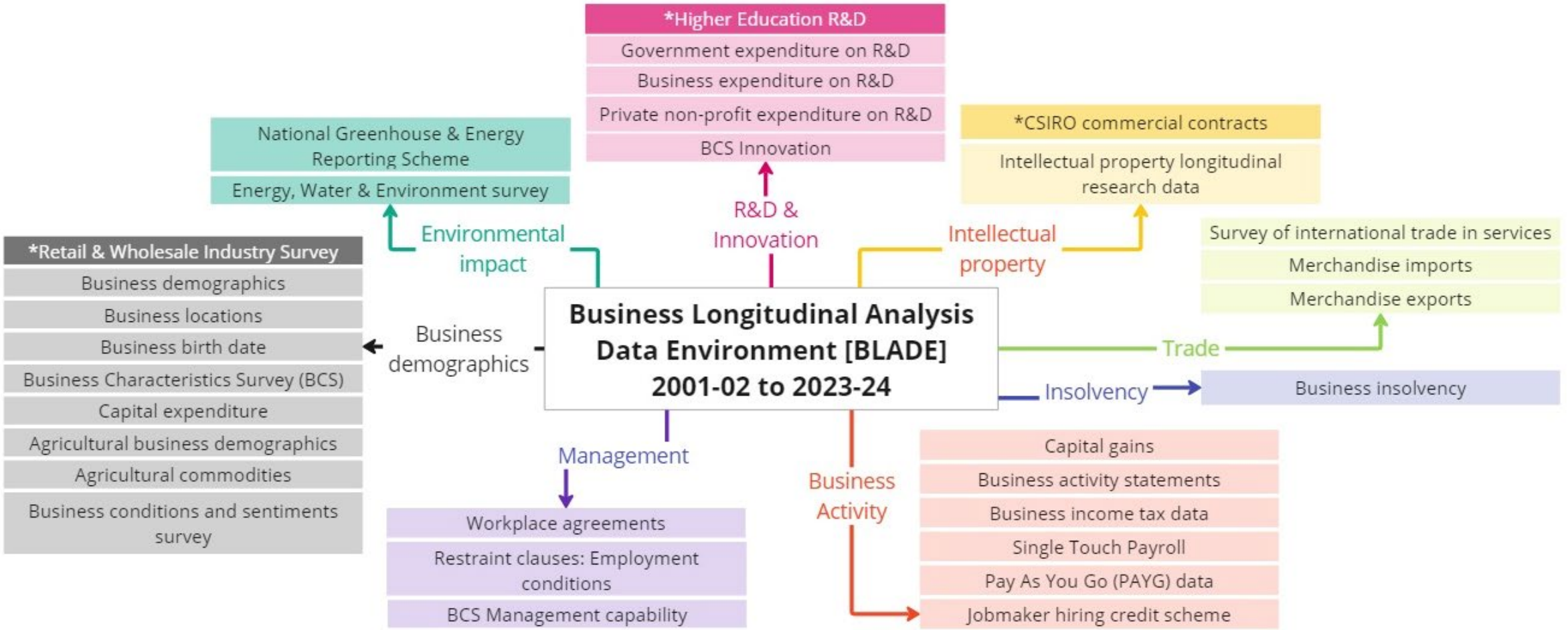
- Retail Trade (mid-2025)
- National Accounts measures of household consumption are shifting from using survey data to private sector big data
- Using supermarket scanner data to estimate food consumption in National Accounts
- Modernising agricultural statistics – using data based on levies collected by government, satellite imagery data, biosecurity data, data held by industry bodies and collected by farm machinery and farm software. Reduces burden on farmers by 97.5%
- Accounting Software project – survey responses for Quarterly Business Indicator Survey are uploaded digitally from small businesses' accounting software packages

Replace select data items in existing surveys

- Replace medications survey data in the National Health Survey with admin data by linking the National Health Survey to Pharmaceutical Benefits Scheme.
- Replace superannuation survey data for the Survey of Income and Housing with admin data from the ATO by linking the Survey of Income and Housing to ATO data.
  - This was unsuccessful due to insufficient coverage (excludes youth/migrants) and unacceptable lags.

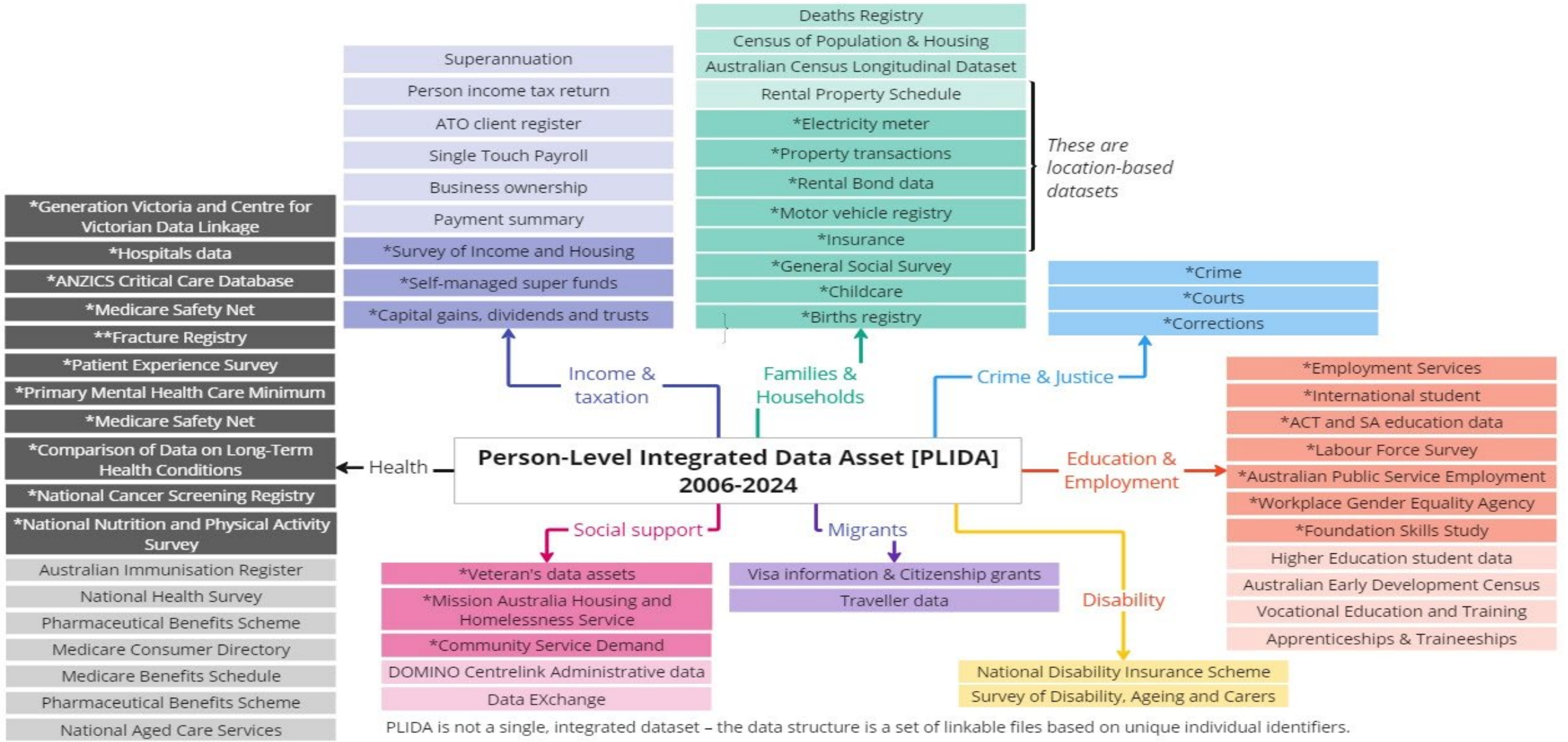
Assist with sample selection and survey response follow-up

- Supplementing the address register with information from the Person Level Integrated Data Asset (PLIDA) to improve sample selection and the representation of small population groups that are important to surveys.
- The 2024 Time Use Survey is a digital-only collection. Admin data was used to over-sample households in areas where response rates have been low because they are hard-to-reach with a digital collection. A likelihood of receiving a digital response was assigned to each household using a model combining PLIDA characteristics with survey response information from our Labour Force Survey.
  - This approach also provided the opportunity to use PLIDA information during survey weighting and estimation to further improve the quality of final survey results.
  - We will expand this approach for upcoming surveys where we wish to sample households with specific characteristics (e.g., pensioners).
  - Managing response bias in our household surveys will be important as we rely less on interviewers and more on digital collection.
- We are developing an enriched address register to enable us to target follow-up effort to non-responding households which have the greatest statistical value for a particular collection – to improve return on investment.



BLADE is not a single, integrated dataset – the data structure is a set of linkable files based on unique individual identifiers.

*An * indicates that the data supply is in negotiation or data integration is underway.*



PLIDA is not a single, integrated dataset – the data structure is a set of linkable files based on unique individual identifiers.

*An * indicates that the data supply is in negotiation or data integration is underway.*

ACT and SA data are being linked to PLIDA for the purposes of the Life Course Data Initiative and access will be limited to approved LCDI projects during the pilot phase.