



A-PLIDA-analysis: Using PLIDA for public policy research and reporting

Research paper

The Productivity Commission acknowledges the Traditional Owners of Country throughout Australia and their continuing connection to land, waters and community. We pay our respects to their Cultures, Country and Elders past and present.

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information can be obtained from the Commission's website: www.pc.gov.au.

© Commonwealth of Australia 2024



With the exception of the Commonwealth Coat of Arms and content supplied by third parties, this copyright work is licensed under a Creative Commons Attribution 4.0 International licence. In essence, you are free to copy, communicate and adapt the work, as long as you attribute the work to the Commission (but not in any way that suggests the Commission endorses you or your use) and abide by the other licence terms. The licence can be viewed at: <https://creativecommons.org/licenses/by/4.0>.

The terms under which the Coat of Arms can be used are detailed at: www.pmc.gov.au/government/commonwealth-coat-arms.

Wherever a third party holds copyright in this material the copyright remains with that party. Their permission may be required to use the material, please contact them directly.

ISBN 978-1-74037-807-9 (online)

An appropriate reference for this publication is:
Productivity Commission 2024, *A-PLIDA-nalysis: Using PLIDA for public policy research and reporting*, Research paper, Canberra

Publication enquiries:
Phone 03 9653 2244 | Email publications@pc.gov.au

Contents

Acknowledgements	iv
Key points	1
1. Administrative data is a growing evidence base for public policy analysis	2
2. PLIDA offers important insights into a range of policy areas	7
3. Understanding PLIDA is essential for productive use	19
4. Navigating some of PLIDA's limitations	24
5. Ongoing development continues to support better access and greater use	31
Appendix A – PLIDA data items	35
Appendix B – Who has produced what?	38
Abbreviations	41
References	43

Acknowledgements

The Productivity Commission (PC) thanks and acknowledges the valuable assistance and feedback from members of the Australian Bureau of Statistics (ABS), particularly from Dr Luke Hendrickson, Rachel Price, Jessica Solla, Kat Walsh, and Sybille McKeown. The PC also thanks for feedback to early drafts of the paper from experts including Dr Nathan Deutscher, Carrie Samuels, Professor Jennifer Welsh, Professor Nicholas Biddle, Jonathan Hambur, and Dr Peter Rankin.

This paper uses data supplied to the ABS under the *Taxation Administration Act 1953*, *A New Tax System (Australian Business Number) Act 1999*, *Australian Border Force Act 2015*, *Social Security (Administration) Act 1999*, *A New Tax System (Family Assistance) (Administration) Act 1999*, *Paid Parental Leave Act 2010* and/or the *Student Assistance Act 1973*. Such data may only be used for the purpose of administering the *Census and Statistics Act 1905* or performance of functions of the ABS as set out in section 6 of the *Australian Bureau of Statistics (ABS) Act 1975*. No individual information collected under the *Census and Statistics Act 1905* is provided back to custodians for administrative or regulatory purposes. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes and is not related to the ability of the data to support the Australian Taxation Office, Australian Business Register, Department of Social Services and/or Department of Home Affairs' core operational requirements.

Legislative requirements to ensure privacy and secrecy of these data have been followed. For access to PLIDA and/or BLADE data under section 16A of the ABS Act or enabled by section 15 of the *Census and Statistics (Information Release and Access) Determination 2018*, source data are de-identified and so data about specific individuals has not been viewed in conducting this analysis. In accordance with the *Census and Statistics Act 1905*, results have been treated where necessary to ensure that they are not likely to enable identification of a particular person or organisation.

Finally, the PC acknowledges the staff who worked on this paper. The work was overseen by Commissioner Catherine de Fontenay and Assistant Commissioner Benjamin Mitra-Kahn. The team included Ariun-Erdene Bayarjargal, Ben Cochrane and Cameron Eren. The PC also thanks other staff who provided input and comments on the paper, particularly the Closing the Gap, Report on Government Services, and Health and Ageing teams.

Key points

- * **Linked administrative data has become an important resource for public policy research, offering insights into how policy settings can impact individual, household, and company behaviour.**
- * **The Person Level Integrated Data Asset (PLIDA) is a linked administrative data asset managed by the Australian Bureau of Statistics (ABS). PLIDA provides secure access to deidentified data on the Australian population across a range public policy areas like health, education, employment, taxation, government transfer payments, and population demographics.**
- * **PLIDA presents a number of opportunities for the research and reporting functions of the Productivity Commission (PC). Additional opportunities can become available by linking PLIDA with other large administrative and survey datasets, such as the ABS's business-focused data asset, the Business Longitudinal Analysis Data Environment (BLADE).**
- * **Recent directions towards inclusion of state and territory data will further increase the range of research possibilities offered by PLIDA.**
- * **Ongoing efforts to increase code sharing, and the introduction of *Core modules*, promise to lower entry costs for new PLIDA users, encouraging greater use of the data asset over time.**

1. Administrative data is a growing evidence base for public policy analysis

People generate data at an incredible scale, creating a rich trove of information that has the potential to transform the way we explore and respond to the policy challenges of our time. Every payslip we receive, every prescription we pick-up, every doctor's appointment we attend, tax return we submit, or childcare session we book, all create unit records in government systems. These government records, known as administrative data, have the potential to be joined together to enable analysis of public policy problems for which a single dataset would be insufficient (Gruen 2024). For researchers, this deidentified and secure data presents an opportunity to move from small surveys with several thousand respondents, to multi-million anonymised datasets – the equivalent of moving from single track audio to surround-sound when listening to the stories that society is telling policy makers.

Australian linked data assets have matured to the stage where data assets are close to covering the entire population of Australia, and often over a number of years. Importantly, these data assets are anonymised and heavily protected, meaning that while some administrative data assets cover close to the entire population of Australia, the identities of individuals in the data asset remain protected.

By linking administrative data together, public policy researchers can gain potentially important insights into individual and household behaviour. As one example amongst many, linking administrative data like tax return data with government payments, and childcare data might provide some insight into the relationship between income levels and childcare decisions. This analysis can be further enriched by linking to survey data like the Labour Force Survey, which provides additional detail on labour force decisions.

Given that statistical inferences do not identify strictly causal relationships, the underlying logic between relationships identified in administrative data can then be tested further with surveys of individuals in the observed population cohort, or randomised control trials that focus on the specific behavioural patterns observed in the administrative data. This provides an opportunity to advance our understanding of human behaviour and social phenomena, and to thereby improve the effectiveness of government services and public policy (Connelly et al. 2016; Curchin and Edwards 2021; Hindmarsh et al. 2021; Leigh 2024).

Researchers are being increasingly drawn to Australia's rich administrative data assets, similar to the way in which microsimulation specialists were drawn to the data-rich Nordic countries in earlier years (Gruen and Goldbloom 2008, p.18).

Despite this growth, use of administrative data is still in its relative infancy in Australia. As of November 2024, there are 2,664 active users in the DataLab, the platform through which researchers can access administrative data assets and microdata in Australia. Of these active users of DataLab, 1,996 are analysts who use the data, and 668 are discussants who review analysts' work. As of 30 September 2024, there were 305 active projects accessing integrated data through DataLab. Of these, 163 projects use the Person Level Integrated Data Asset (PLIDA), and 73 use the linked PLIDA and the Business Longitudinal Analysis Data Environment (BLADE), a business-focused data asset. Despite growth in the use of PLIDA, visibility of completed research outputs is limited, partially limiting the traceability of outputs (section 5).

This paper aims to identify opportunities for the use of PLIDA in PC research and reporting, as well as to build understanding of research opportunities within the broader Australian public policy community, and to

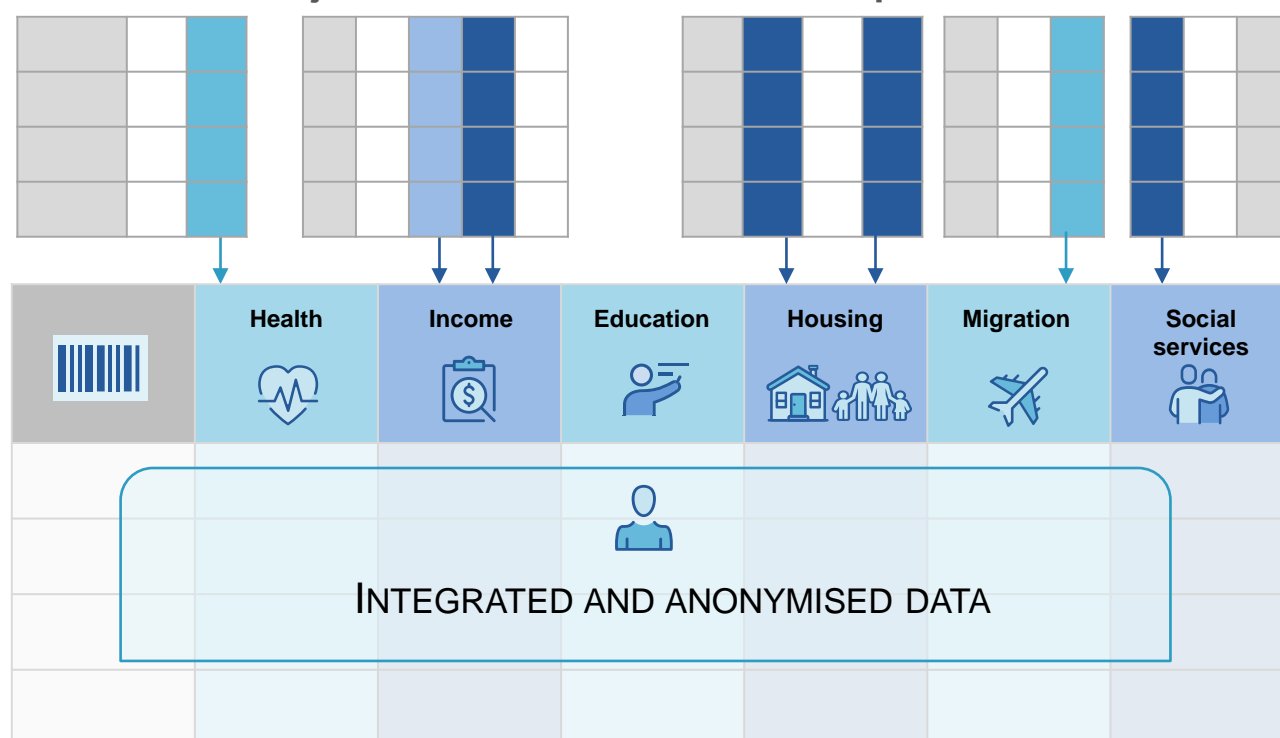
identify development directions for the data asset that are relevant to both new and existing researchers. It does so from a user perspective.

Linked administrative data is powerful

The power of linked administrative data comes from its size and regularity. Data can be pieced together to provide coverage across a large population, often over a number of years, enabling extraordinarily large sample sizes for policy research with no additional cost of collecting the information (figure 1.1). In contrast to surveys, administrative data is less likely to suffer from specific individuals dropping out of the survey over time (known as survey attrition), non-responses or survey bias. The can also provide larger sample sizes for previously less visible sections of the population or those who may be hard to get into contact with, such as geographically remote communities and individuals who are not interested in participating in studies and surveys (PC 2018, p. 11). This potential was reflected in the PC's 1999 inquiry on gambling where the majority of surveyed respondents indicated that without counselling they would not have answered a survey about gambling (PC 1999).

Figure 1.1 – Linking administrative data creates incredibly useful assets

De-identified data is joined across sources to create a rich picture



However, administrative data is not without its challenges (table 1.1). Some of these challenges arise from the fact that administrative data is not collected for research purposes. This means that it can be in a messy format and can be of a computationally demanding size. In addition, certain groups may be over- or under-represented in the data because some groups use government services more than others, which may lead to biases (Shaw et al. 2022). There might also be instances where what is being captured by administrative data is not the most accurate reflection of underlying circumstances. For example, increases in the tax free threshold over time can change the overall proportion of income reported to the Australian Tax Office (ATO) over that time period. Income data like Single Touch Payroll (STP) data captures weekly, fortnightly or monthly earnings, but not hours of work

or the standard hourly wage rate of an individual. In addition, tax incentives and the risks of providing inaccurate information to tax authorities, might affect what is reported to the ATO. The sheer scale of the data can also make certain types of analysis challenging, presenting a steep learning curve to new administrative data researchers. Both data custodians and researchers have a role to play in managing these challenges, whether through custodians developing higher quality datasets and detailed metadata, or researchers sharing code and practices to work with administrative data.

Table 1.1 – Advantages and disadvantages of using (linked) administrative data for research

Advantages	Disadvantages
Large coverage of population	Not built explicitly for research purposes, variables and recorded data may have interpretation or quality issue
Routinely updated	Linking across datasets can be difficult or imperfect
Limited costs required for use (already collected for an administrative purpose)	Incompleteness, program and system changes, and disruptions can complicate historical comparisons
Data often extends historically, making it possible to create panel datasets	Ongoing access to data is reliant on data providers support and clearance, may be treated as a lower priority
Captures small populations that may not normally be captured in surveys (rural or remote, or susceptible to survey attrition)	Large data can present computational hurdles and challenges
Can be more accurate, and less subjective than survey data	Can have higher upfront costs for researchers

Source: Harron et al. (2017); Smith et al. (2004).

Globally, the use of administrative microdata for research has been accelerating (Card et al. 2010; Cole et al. 2021; UNSD 2020). Integrating and linking administrative data across government agencies has become an essential tool for research and policy analysis in many countries. Nordic countries (Denmark, Sweden, Norway, and Finland) are decades into developing and using linked administrative data, where unique person identity numbers enable them to link a range of individual-level information across a number of government-maintained nationwide registries (Jervelund and De Montgomery 2020; UNSD 2020). With the development of PLIDA, Australia is joining these global leaders in linked administrative data space.

PLIDA is one of Australia’s key data assets

PLIDA, formerly known as the Multi-Agency Data Integration Project (MADIP),¹ links the person-level administrative and survey information collected by various government departments and agencies. MADIP commenced in 2015, and first became available to researchers as the MADIP Basic Longitudinal Extract (BLE) 2011–2016 microdata product in 2018.² The Australian Bureau of Statistics (ABS), as the Accredited Integrating Authority³ for PLIDA, collects data from a range of data custodians,⁴ combining datasets to provide

¹ MADIP (Multi-Agency Data Integration Project) was renamed to PLIDA in 2023, to better reflect the nature of the data asset itself, rather than the process through which it was created.

² Details can be found [1700.0 - Microdata: Multi-Agency Data Integration Project, Australia](#).

³ The ABS is the Accredited Integrating Authority under the Commonwealth Data Integration Interim Arrangements since 2012. In the new DATA Scheme, the ABS is an Accredited Data Service Provider for complex data integration, secure data access services, and de-identification. [Accredited Entity Register | Office of the National Data Commissioner](#).

⁴ The data custodians include the Australian Taxation Office, Services Australia, the Department of Social Services, the Department of Health and Aged Care and others. A complete list of data custodians can be found at [PLIDA data and legislation | Australian Bureau of Statistics \(abs.gov.au\)](#).

detailed information on the way in which the Australian population engages with social, health, and education services, as well as their incomes, the government payments they receive, and the tax that they pay.

PLIDA is protected by a number of security and privacy controls (box 1.1). Strict adherence to the Separation Principle⁵ and application of the global practice of the Five Safes Framework,⁶ direct identifiers are removed from data, PLIDA is accessed through a secure DataLab environment which requires users to be authorised and authenticated, and projects and outputs are subject to rigorous assessment and approval processes (ABS 2021a; Parker 2017).⁷

Box 1.1 – PLIDA is protected by a number of security and privacy controls

The Australian Bureau of Statistics (ABS) ensures the privacy of personal information in the Person-Level Integrated Data Asset (PLIDA) is secure, protected, and researchers access the data in a safe environment. The ABS adheres to legislative requirements, conducts privacy impact assessments, and applies governance processes and accountabilities (ABS 2024f).

- The strategic direction and oversight of PLIDA is governed by the PLIDA Board. The Board makes decisions about data acquisition and use, linkage infrastructure, risk and issue management, governance and legal advice, and stakeholder engagement and communication strategies.
- The *Privacy Act 1998* (Privacy Act) and the Australian Privacy Principles (APP) are legislative requirements for collecting and using personal information. Specific data handling practices for PLIDA are outlined in the PLIDA Privacy Statement.
- The ABS conducts Privacy Impact Assessments (PIA) for PLIDA on behalf of the PLIDA Board. PIAs assess proposed updates such as the addition of new datasets to PLIDA and provides impact assessments of these updates including whether they affect the ABS's compliance with the Privacy Act, APP, and other relevant privacy related policies.
- In addition to legislative and privacy protections, ABS adherence to a range of data handling principles and standards ensures that integrated data is safe and secure. These include, but are not limited to:
 - Five Safes Framework
 - High Level Principles for Data Integration
 - Separation Principle
 - Data minimisation principles⁸
 - Transparency measures and data handling practices

Further details on each of these principles and practices can be found in the ABS website.⁹

Source: ABS (2024a, 2024e).

Since its establishment, PLIDA has expanded to include a broad range of administrative data and surveys around eight themes (figure 1.2). The end result is a rich data asset for public policy research that continues to expand.

⁵ The Separation Principle keeps personal identifier (or linkage) information separately from analytical information, enabling no individual or business is identifiable in the data. [Keeping integrated data safe | Australian Bureau of Statistics](#).

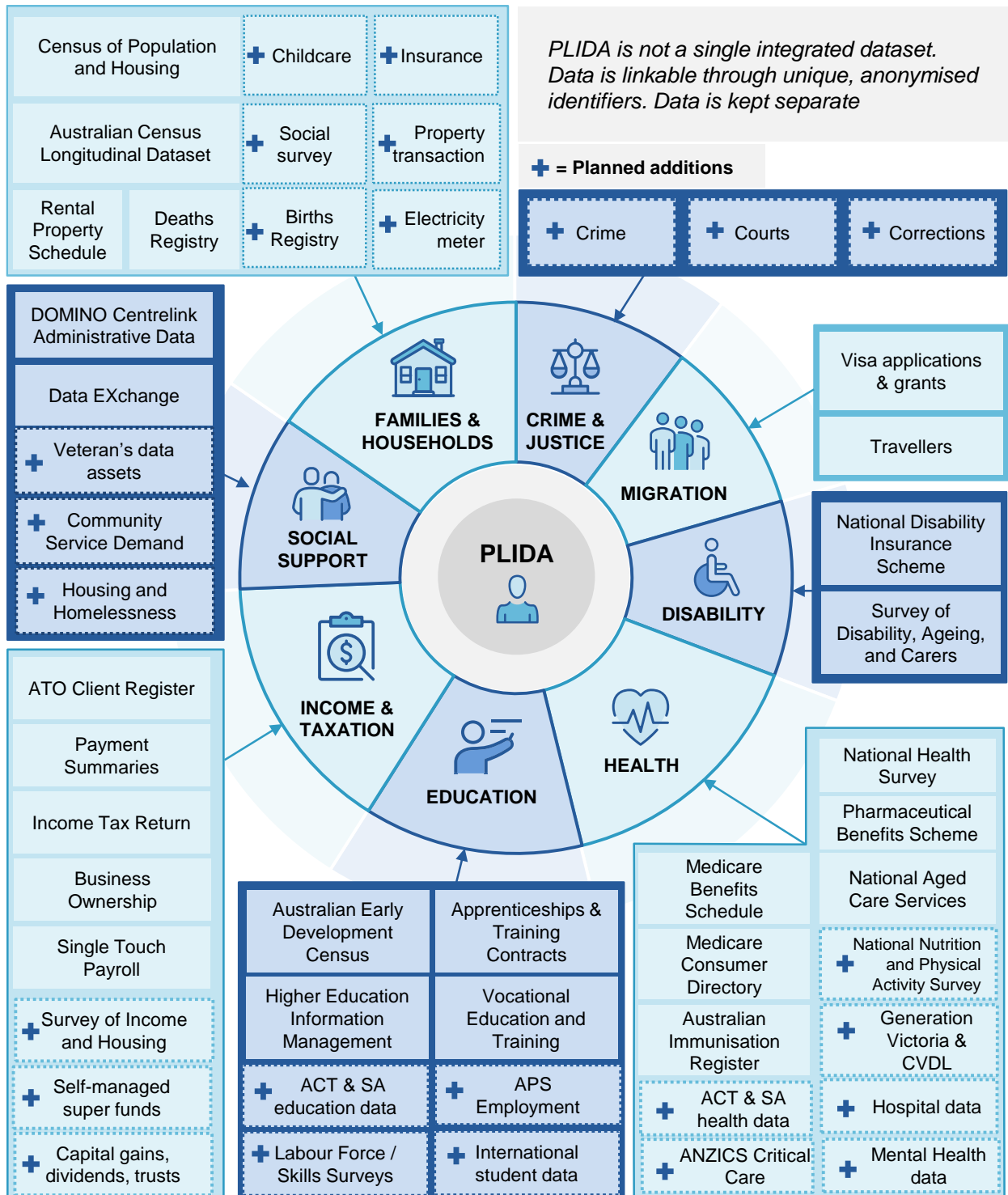
⁶ The Five Safes Framework is an internationally recognized risk management model, helping make decisions about effective use of data which is confidential or sensitive. [Five Safes framework | Australian Bureau of Statistics \(abs.gov.au\)](#).

⁷ PLIDA is accessible only to authorised users who have completed safe-researcher trainings.

⁸ Data minimisation principles ensure that researchers only gain access to data that is necessary for their research purposes.

⁹ See: [Keeping integrated data safe | Australian Bureau of Statistics](#), [Privacy in PLIDA | Australian Bureau of Statistics](#).

Figure 1.2 – PLIDA datasets by themes^{a,b}



a. DOMINO = Data Over Multiple Individual Occurrences – Centrelink Administrative Data (DOMINO-CAD); CVDL = Centre for Victorian Data Linkage; ANZICS = Australian and New Zealand Intensive Care Society; APS = Australian Public Service **b.** Not all of these datasets are available in the modular product, given some are linked for specific projects and are under restrictive data supply agreements. The current PLIDA Data Item List is accessible via [Microdata: Person Level Integrated Data Asset \(PLIDA\)](https://microdata.abs.gov.au) | Australian Bureau of Statistics (abs.gov.au).

Sources: ABS (2024b) and ABS, personal communications, 11 November 2024.

PLIDA is provided to approved researchers as a collection of modules, PLIDA Modular Product (PMP), where end-user products are created from the datasets within the DataLab. Modules can differ depending on the datasets provided by data custodians, often split by themes and by time periods. As of September 2024, PMP contains 37 datasets (figure 3.3), 270 modules and 42,033 variables (11,383 unique variables).¹⁰

2. PLIDA offers important insights into a range of policy areas

The research opportunities presented by PLIDA are extensive and are starting to be realised. As of October 2024, there were 356 listed PLIDA projects across Australian Government, state and territory governments, and academic and private research institutions.¹¹

The ABS publishes project summaries which are provided by researcher teams. The earliest project listed commenced in 2017. Project summaries indicate that a large amount of work has so far focused on assessing the impact of existing policy settings on health, ageing, disability, education, income, and labour market outcomes (figure 2.1). Health analysis is particularly common, likely because health researchers were early users of the linked administrative data in PLIDA and its predecessor, MADIP.

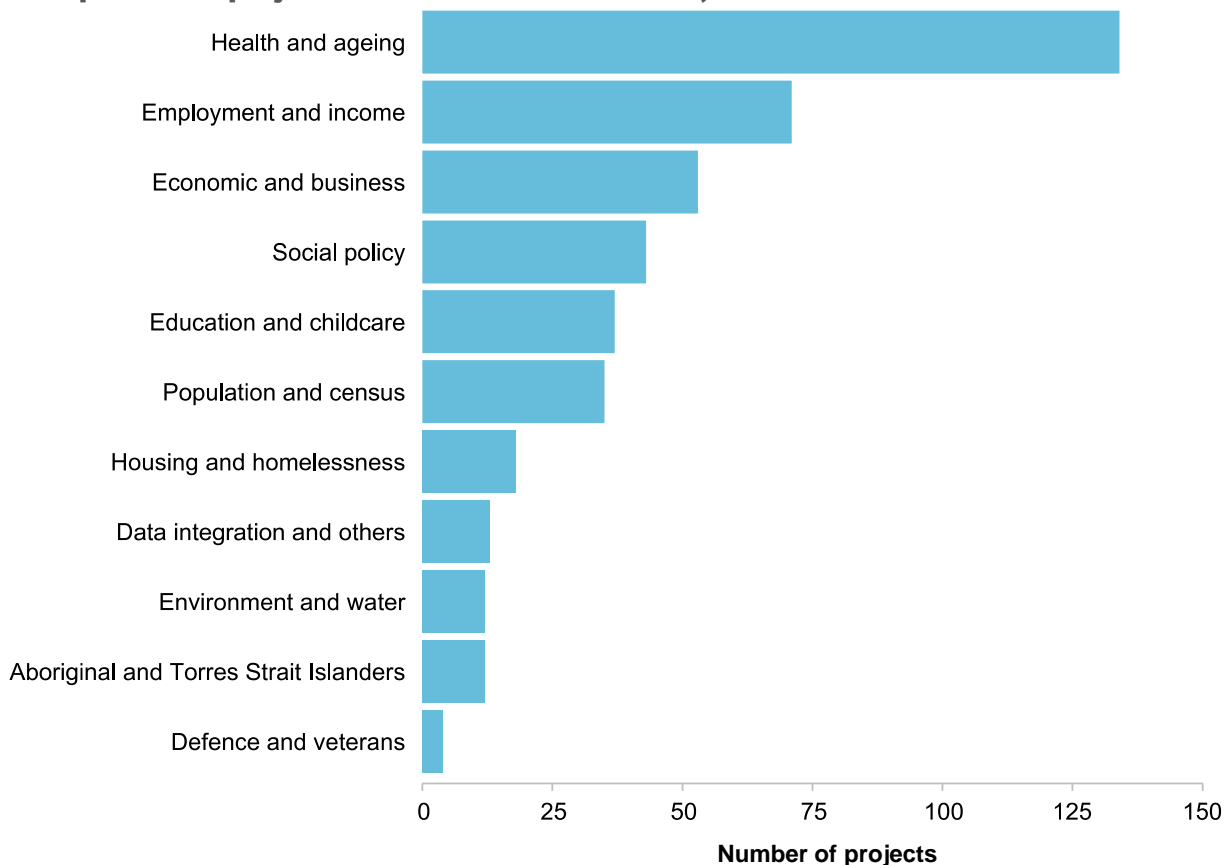
The breadth of research possibilities reflects PLIDA's holding of information about individuals across a range of domains such as education, taxation, income, government payments, health, and demographics. Importantly, data can be combined *across* these domains to provide important insights that are relevant to policy makers. Some of research opportunities are summarised below.

¹⁰ The details of the datasets are being updated on an ongoing basis and can be found at [Microdata: Person Level Integrated Data Asset \(PLIDA\) | Australian Bureau of Statistics \(abs.gov.au\)](#).

¹¹ This number differs from the ABS website PLIDA project register which is updated quarterly.

Figure 2.1 – ABS PLIDA projects by categories^a

ABS published project summaries 2017 to 2024, as of October 2024

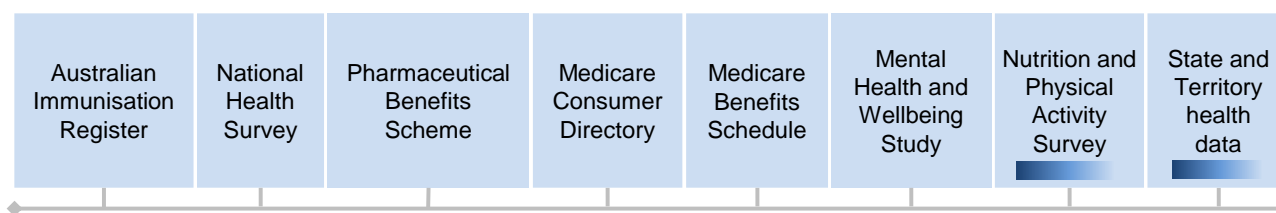


a. PC assigned categories based on project summaries. A project may fall into multiple categories (for example, education *and* income) and therefore total project count may be greater than total project count.

Source: PC calculations based on ABS (2024c).

Health

PLIDA provides insight into various aspects of the health care system in Australia, covering the datasets of the Medicare Benefits Schedule (MBS), the Pharmaceutical Benefits Scheme (PBS), the National Health Survey (NHS), Mental Health and Wellbeing Survey (MHWS) and Australian Immunisation Records (AIR). PLIDA continues to expand with the integration of additional datasets in the health domain such as National Nutrition and Physical Activity Survey, and selected state and territory datasets (figure 2.2). This coverage enables research into the use of health services in Australia and the health and wellbeing outcomes of Australians. For example, adding data from the AIR to PLIDA provided insight into the socio-economic determinants of vaccine uptake by various sub-groups of the population, and allowed for assessments of the effectiveness of vaccines across the broader population (AIHW 2022, 2024b; Biddle et al. 2022; Liu et al. 2023). Importantly, the findings from these linked data helped inform policy interventions for those most at risk and helped to understand causes of deaths among particular population groups such as aged care residents (Welsh et al. 2024). Linkage also enabled the identification of groups with low-vaccine uptake in groups who speak languages other than English, which helped inform communication and outreach approaches for these groups (Gruen 2024).

Figure 2.2 – PLIDA’s health-related datasets

The PC used PLIDA (MADIP) in the Mental Health inquiry, examining mental health through geographic, educational and employment lenses (PC 2020). The PC is currently using PLIDA to develop its understanding of private health insurance and hospital service use in Australia (box 2.1).

PLIDA is also proving useful for some reporting purposes. The AIR-PLIDA linkage, together with provisional mortality data from the ABS, has improved reporting on rates of vaccination and deaths among selected population groups such as residential aged care residents and National Disability Insurance Scheme recipients (AIHW 2024a; DoHAC 2024).

Box 2.1 – Using PLIDA for Health policy research at PC

The PC’s Health and Ageing policy stream is using PLIDA for its current research *Takeup and use of private health insurance by young people*. It will use tax return data to identify individual’s private health insurance status, including premiums paid and coverage levels. The analysis will also cover how individuals change their coverage and policies over time.

PLIDA has enabled the project team to observe individuals over time and produce demographic analysis that would otherwise be unachievable. In the context of examining the private health system, it has been valuable to explore the payment and tax levers that are associated with insurance takeup and use, especially amongst younger people.

Nevertheless, PLIDA has limitations for our analysis. The time series available is shorter compared to other data sources such as the ATO Longitudinal Information Files (ALIFE). In addition, PLIDA is not linked to key public health services data that is relevant to the Health and Ageing policy stream’s research.

Integrating PLIDA with public health services datasets is vital to support a system-wide, person-centric perspective of Australia’s health system. For instance, the Australian Institute of Health and Welfare (AIHW) maintains comprehensive hospital admissions data at the patient level that could be integrated with PLIDA. The ability to combine healthcare data with patients’ tax and welfare data could support better policy development aimed at improving health outcomes and system efficiency. The ABS is working towards developing an expanded linkage of health data, with a recent Privacy Impact Assessment being conducted for adding health data to PLIDA (ABS 2024d).

Researchers have been able to generate important policy insights by linking some state and territory data with PLIDA. For example, the one-off linkage of New South Wales Cancer Registry with PLIDA enabled researchers to estimate the likelihood of being diagnosed with lung cancer by socio-demographic characteristics. Such insights could be used to more effectively target cancer control and detection programs (Banham et al. 2024).

Similarly, the incorporation of state and territory Deaths Register data into PLIDA has also allowed researchers to explore the relationship between life expectancy and educational attainment levels (Korda et

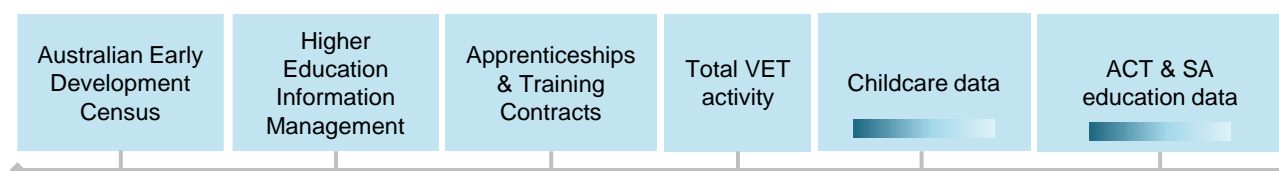
al. 2020). This research could be built upon using the rich sociodemographic and geographic data in PLIDA to further understand the determinants of individual health outcomes.

National health-related research is limited by the availability of state and territory health data. Planned linkage with PLIDA is likely to support a more holistic and integrated view of the Australian health system over time, but gaps currently exist. Improvements to integrated data was also identified as important by the COVID-19 Response Inquiry for supporting the development of effective policy responses to national health emergencies (Department of the Prime Minister and Cabinet and Commonwealth of Australia 2024). The status of state and territory data integration into PLIDA more generally is explored further in section 5.

Education

PLIDA has become a valuable resource for early and post-school education research. Education-related datasets include the Australian Early Development Census (AEDC), which measures the development and readiness of children for school; Apprentice and Trainee data, which tracks the progress and outcomes of apprentices and trainees in various industries and occupations; Total Vocational Education and Training (VET) activity data, which contains information about individuals who enrolled in and completed recognised VET courses; and the Higher Education data, which provides information on the enrolments, completions, student characteristics and attrition rates of higher education students (figure 2.3). Selected state and territory education data and Commonwealth Childcare Subsidy Management data are expected additions that would enable further research possibilities in the education domain (figure 1.2 and table 5.1).

Figure 2.3 – PLIDA’s education-related datasets



The integration of these datasets into PLIDA has offered important insights into the relationship between education and a broader range of indicators. For example, PLIDA has allowed for examination of the relationship between the quality of early childhood education and care that children have access to and their developmental outcomes across each of the AEDC domains, by various socio-demographic characteristics (Rankin et al. 2024; Tang et al. 2024). These findings could be informative for improving the guidance on development milestones in the government’s Early Years Learning Framework and the National Quality Standards for service providers.

PLIDA’s rich longitudinal datasets provide information on children’s health and education outcomes that can be observed with respect to their family’s socio-demographic status. This enables researchers to assess the relationship between indicators of childhood disadvantage and the achievement of developmental milestones during childhood. Such findings can help researchers develop a greater appreciation of the potentially complex nature of disadvantage (Gray et al. 2023; Pham et al. 2024).

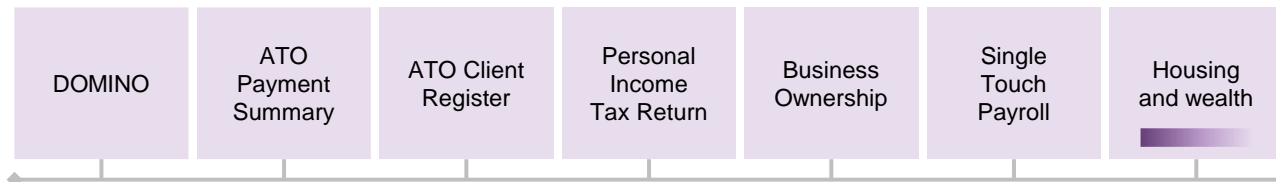
VET enrolment and Higher Education data in PLIDA has enabled reporting on various indicators of educational outcomes for Aboriginal and Torres Strait Islander people (Nicolaou et al. 2023). Some researchers have gone further, studying the relationship between Aboriginal and Torres Strait Islander participation in vocational and education training and their longer-term employment outcomes (NCVER 2023). Policy makers have explored these insights to determine which VET courses might warrant greater government support than others (DESE 2020).

As major providers of government services in Australia, incorporating State and Territory data into PLIDA can provide a clearer picture of the nature of disadvantage in Australia. Perhaps for this reason, the ABS is currently supporting the incorporation of selected State and Territory data into PLIDA. In time, this might help to support the more effective targeting of government services towards entrenched disadvantage in some communities (section 5).

Income, employment, and taxation

PLIDA contains several datasets that include detailed information on the income levels and employment status of Australians (figure 2.4). These include Personal Income Tax (PIT) data, which provides the income levels of all individuals who lodge a tax return, along with their respective deductions, tax offsets and tax liabilities. PLIDA also includes STP data, which includes earnings paid by businesses to their employees. These datasets can potentially provide researchers with insight into the impacts of the tax system and transfer system on the labour market decisions of workers. They also allow researchers to explore the broader range of factors that potentially shape employment outcomes and income levels across regions, industries and occupations. For example, by using transfer payment details provided by the DOMINO Centrelink Administrative Data (DOMINO) linkage within PLIDA, researchers have been able to build an understanding of the individual and family characteristics of parental leave payment recipients over time (Baxter and Budinski 2023).

Figure 2.4 – PLIDA’s income, employment and taxation datasets



STP data in PLIDA enables researchers to use an alternative approach for quality-adjusted labour productivity (Bruno et al. 2024). STP is high frequency and was used in the Commonwealth Treasury’s Labour Market Tracker. During the COVID-19 pandemic, this was updated with a lag of only 3-4 weeks, which enabled timely data on indicators of employment and income (Hambur et al. 2022). Now as an ongoing linked dataset in PLIDA, STP is updated monthly, providing users with relatively frequent data.

PLIDA allows for the integration of other administrative data on income and employment, such as transfer system payments, superannuation contributions and employer-reported earnings (Hamilton et al. 2023; PC 2024a). Integrating these provides the opportunity to combine different sources of income data (personal income tax, transfer payments, survey and Census) which enables the creation or imputation of a more complete income spectrum (Biddle and Marasinghe 2021). The PC used multiple income data sources in PLIDA to construct income measures for inequality and poverty analysis. These indicators can continually be updated to inform policy related to entrenched disadvantage in the future (box 2.2).

Box 2.2 – Using PLIDA for inequality and poverty indicators

The Productivity Commission (PC) used PLIDA in its 2024 research on inequality and income mobility to measure inequality levels and to produce poverty indicators. These indicators have traditionally been estimated using survey data such as the ABS Household Expenditure Survey (HES) and Survey of Income and Housing (SIH), and the Household Income and Labour Dynamics in Australia (HILDA) (Biddle et al. 2019; PC 2018). These survey data have been subject to changes in definitions of income and measurement, however. For example, the introduction of computer-assisted interviewing to SIH increased income reporting, particularly by high-income earners, which led to higher Gini coefficient estimates (PC 2018). Shifting coverage of the population through non-random survey attrition for HILDA also affects comparability of estimates over time.

PLIDA provides access to a range of personal income data

Integrated personal income tax data in PLIDA can allow for more accurate estimates of inequality and poverty indicators over time. This is because personal income data reported to the Australian Taxation Office (ATO) has more consistent income definitions over time.

It is not perfect, however. Personal income tax data is generally more skewed towards the top of the income distribution than the survey data, partly reflecting the number of people that do not need to lodge tax returns because they are below Australia’s \$18 200 tax free threshold. Moreover, tax integrity measures that are successful in encouraging individuals to report their income accurately to the ATO and increases in the tax-free threshold over time can change income distribution estimates in a way that is unrelated to actual income changes (section 4).

The table below shows that the average income of the top percentile is significantly higher in the ATO data available through PLIDA than that in the SIH. As a result, the PC generated different estimates of the Gini coefficient and poverty rate when using PLIDA, relative to those estimated using SIH/HES data (PC 2024a).

	SIH/HILDA data	Administrative data
Top 1%’s average weekly income	\$9,708 with standard deviation of	\$12,270 with standard deviation of
(of the population aged 15 and over)	\$5,704 in SIH 2019-20.	\$12,283 in the ATO 2019-20 in PLIDA.

Using ATO data alone will miss individuals who do not file a tax return in a given year (section 4). The PC was able to navigate this issue by linking tax return data with payment summary data issued by employers and lodged with the ATO and government income support payment data from the Department of Social Services. This provided income information for more than 90% of Australians and decreased PC estimates of mean and median income by roughly 20% compared with ATO data only.

But comes with some caveats

Household-equivalised income is a commonly used indicator in inequality and poverty analysis. At the time of publication of the PC’s 2024 inequality and mobility research papers reports the only source of information available in PLIDA to construct household income was the Census that is held every five years. The resulting time gaps in PLIDA limited the estimation of poverty and inequality in any given year. As a result the PC estimated the Gini coefficient for the 2019-20 to 2021-22 time period using 2021 Census household structure (PC 2024a). Not reflecting household dynamics that might have occurred in years for which there is not updated Census data may have resulted in either over- or under-estimation of Australia’s Gini coefficient and poverty rate in those years.

Box 2.2 – Using PLIDA for inequality and poverty indicators

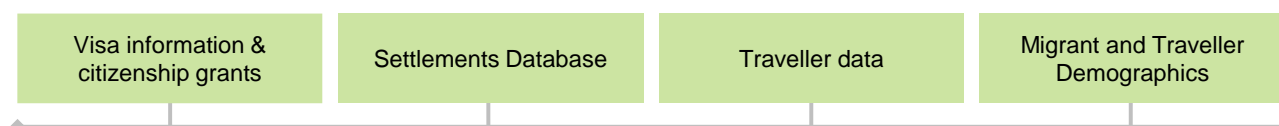
However, PLIDA’s ongoing expansion and deepening provides an opportunity for future inequality, mobility and poverty related analysis, in at least three ways:

1. The series of inequality and poverty indicators can be created for the years Census is available. The planned introduction of family and household compositions in Core modules by the ABS will allow for continuous estimation of these measures.
2. Similarly, life course (intragenerational) and intergenerational mobility analysis can be conducted between two points of time with observed information for generations.
3. Potential integration of wealth data into PLIDA could provide greater insight into inequality and social mobility dynamics in the Australian community.

Migration

PLIDA provides access to several datasets on various aspects of migration, such as the pattern, duration, motivation, and effect of moving within or across countries (figure 2.5). These include the Migrant and Traveller Demographics dataset, which combines data from a range of sources to create a comprehensive picture of migration flows and stocks; the Visa Applications and Grants dataset, which provides information about people who have applied for visas to migrate to Australia; and the Settlements dataset, which provides information on permanent migrants who have settled in Australia. A limitation of PLIDA for migration analysis is the lower coverage of new and temporary migrants in linked data (section 4 for further information). This is an active consideration for the ABS, who continue to support better linkage quality.

Figure 2.5 – PLIDA’s migration-related datasets



These datasets enable researchers to explore the factors that influence migration decisions and outcomes, as well as the impacts of migration on human capital, labour market participation, employment outcomes and fiscal implications (OECD 2023; Varela et al. 2021). Linked to their respective datasets, labour market outcomes such as skill (mis)matches for skilled migrants (Cebulla et al. 2024) and health outcomes of migrants can be assessed (AIHW 2023). Related research findings have already helped to inform some migration policy settings. For example, research on the health outcomes of migrants informed the development of the Refugee and Humanitarian Entrant Settlement and Integration Outcomes Framework (AIHW 2024a). The ABS and the Department of Home Affairs have used PLIDA alongside a range of other data sources to develop enduring indicators for the reporting of migrant data, which can be used to test outcomes against the framework (ABS 2021b).

Integration of datasets can help with the previously difficult task of tracking migrant outcomes in Australia. Doing so can also present broader geographical insights, such as whether permanent migrants on different visa categories become concentrated in certain locations over time. Some research has also explored the feasibility of PLIDA for analysis of internal migration – the movement of residents within Australia – which has concluded that PLIDA is not yet suitable to produce robust population-level migration estimates (Bernard et al. 2024; Laukova et al. 2022).

Labour productivity

The ABS has previously pointed to the potential to use PLIDA for productivity analysis (ABS 2022, p.12). Some of these opportunities are beginning to be realised, particularly where PLIDA and BLADE¹² have been linked with one another (box 2.3). The incorporation of person-level data alongside firm-level information presents the opportunity to gain insights into labour productivity and its drivers.

Box 2.3 – The research possibilities of linking PLIDA and BLADE

Labour market decisions and labour mobility – how and why people move between different jobs and locations for work – have important implications for labour productivity (Andrews and Hansell 2019; Deutscher 2019). When linked to BLADE, PLIDA can offer insights into the potential effects of policy settings on these labour market decisions. For example, this linked data product provides a data source for longitudinal labour productivity and labour mobility analysis that is longer than the Longitudinal Labour Force Survey (LLFS), the existing source for labour mobility analysis, where individuals are followed for only up to eight months.

PLIDA-BLADE linked data can also be a useful data source for observing the effect of changes to the tax and transfer system on labour supply decisions. For example, examining how changes to personal income tax rates and income tax brackets impact the incidence of bracket creep and the extent of labour market participation by different cohorts (PBO 2021; Treasury 2024). Similarly, a PLIDA-BLADE linkage could also be used to assess the effect of changes to government payments on labour market decisions.

Additionally, the PLIDA-BLADE linkage presents opportunities for understanding employment and human capital-related elements of innovation, dynamism and firm growth (Borland and Smedes 2024; Hendrickson et al. 2015). It does by linking employees with employers across a range of individual characteristics.

The PC is currently undertaking research into the rise and subsequent fall in labour productivity before and after COVID-19. Future analysis could use a PLIDA-BLADE linkage to better understand the dynamics that sat behind this phenomenon, by exploring the extent to which labour reallocation, and labour turnover rates, impact the productivity performance of firms.

Linking PLIDA with other ABS surveys could create additional research opportunities. For example, integration of Longitudinal Labour Force Survey (LLFS) data into PLIDA could enrich labour productivity analysis by incorporating information that is not available in administrative data, such as hours worked (Bruno et al. 2024).

The preceding discussion provides a sense of the breadth of public policy research opportunities presented by PLIDA. It is by no means exhaustive. The names of PLIDA projects listed by the ABS provide a sense of the topics that researchers have used PLIDA to explore to date, as well as some of the motivations of this research. Figure 2.6 suggests that PLIDA-related research to date has commonly focused on assessing the ‘outcomes’ and ‘impact’ of different policy settings which is part of the process for driving better policy outcomes over time.

¹² The Business Longitudinal Analysis Data Environment (BLADE) is a data asset that contains tax, trade, and innovation data on all active businesses from 2001-02 in ABS surveys and administrative data.

Figure 2.6 – Frequency of words in ABS’s PLIDA project summaries^a
 ABS published project summary titles 2017–2024, as of October 2024



a. Plural words taken as singular for the purpose of frequency recording, e.g. ‘impacts’ become ‘impact’.

Source: PC calculations based on ABS (2024b).

This desire to improve policy outcomes has led the PC to use PLIDA in a number of its reports (table 2.1) and to explore how PLIDA might also be used in a range of its reporting functions, such as the Closing the Gap reporting and the Report on Government Services (box 2.4).

Table 2.1 – The PC’s use of PLIDA/MADIP

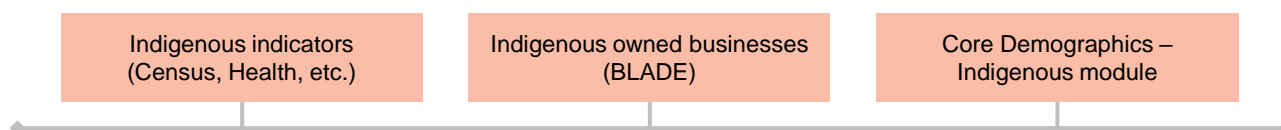
Publication	Description	Year
A snapshot of inequality in Australia	The PC used PLIDA for research into inequality and poverty in Australia and the extent to which they were associated with particular socio-demographic characteristics and geography.	2024
Fairly equal? Economic mobility in Australia		
Early childhood education and care (ECEC) inquiry	The PC gained access to a one-off linkage between SIH MURF and administrative childcare subsidy (CCS) data via PLIDA to estimate the potential effects of ECEC subsidy policy changes on carers’ labour supply decisions.	2024
Mental Health inquiry	The PC used MADIP data for various analyses related to the use and uptake of mental health services. Linked data enabled viewing of mental health outcomes through a geographic, education, and employment lens.	2020
Childcare and Early Childhood Learning inquiry	The 2015 ECEC Inquiry used Social Security and Related Information (SSRI), a precursor dataset of DOMINO which was included in early versions of PLIDA/MADIP to generate insights into the different categories of ECEC customers.	2015

Closing the Gap

Under the National Agreement on Closing the Gap (CtG), the PC is required to report progress against 23 targets and 164 indicators across the Priority Reforms and socio-economic outcomes that impact the life outcomes of Aboriginal and Torres Strait Islander people. Working with the Partnership Working Group, the PC is developing the CtG Information Repository to report against the targets and indicators drawing on a range of data sources. The PC is taking an iterative approach to developing the Information Repository. It currently reports on 42 supporting indicators – about one-quarter of the 164 supporting indicators named in the Agreement.

In addition, governments and the community are working to identify and/or develop data sources for approximately 150 data development items outlined in the CtG Data Development Plan, where there is currently no known data source. PLIDA has the potential to enhance or supplement the range of indicators on the CtGs Information Repository or address known data gaps outlined in the data development plan including for economic indicators, given PLIDA’s linkage with data such as Indigenous-owned businesses in BLADE (figure 2.7). Linkage activities, and the use of linked data is guided by principles for the safe use of Aboriginal and Torres Strait Islander data, discussed below.

Figure 2.7 – Data related to Aboriginal and Torres Strait Islander people



A key issue for the CtG Information Repository is being able to accurately identify Aboriginal and Torres Strait Islander people in available datasets. The current method for doing so relies upon self-identification questions in the Census and other surveys or administrative collections. This approach has reported significant growth in the number of Aboriginal and Torres Strait Islander people, with a significant proportion

of that population growth potentially due to non-demographic factors, such as an increase in the number of people that self-identify as Aboriginal and/or Torres Strait Islander, and/or improvements in information collections processes over recent years (Biddle and Markham 2018; Griffiths et al. 2019).

Another issue is when the numerator and denominator data is sourced from different data collections, causing a numerator–denominator mismatch. This is the case for all indicators that source denominator data from the official ABS population estimates produced following each Census, but the numerator is sourced from independent administrative collections. At times this has resulted in mathematically impossible results (for example, the percentage of children enrolled in preschool is estimated to be greater than 100%).

As a linked data asset, PLIDA could assist with these challenges, by helping to identify people who are identified as Aboriginal and Torres Strait Islander people in one dataset but not another, and to provide a consistent numerator and denominator source. However, identification of Indigenous status with administrative data is a culturally sensitive issue and the ABS is working in partnership with Aboriginal and Torres Strait Islander representative bodies to ensure cultural appropriateness consistent with relevant principles like those embodied in Indigenous Data Sovereignty and the Priority Forms specified in the National Agreement on Closing the Gap.

The ABS is supporting the development of guidelines for users of government administrative data on Aboriginal and Torres Strait Islander people through a multi-year project with the Australian Research Data Commons. This project will be developed in partnership with Aboriginal and Torres Strait Islander people and will take place alongside the ABS's implementation of the APS-wide Framework for Governance of Indigenous Data. The project aims to help researchers better understand Indigenous data, such as the expansive concept of Indigenous data, the context in which they are collected and maintained in PLIDA and the intersection with Indigenous Data Sovereignty and Governance principles.

In addition, for national reporting there needs to be consensus from the Parties (the Coalition of Peaks and Australian, state and territory governments) that any new indicators provide an appropriate standard for assessing progress. For new and innovative datasets, additional time may be required to reach this consensus. Key to this is ensuring Indigenous Data Sovereignty principles are embedded in this process by involving Aboriginal and Torres Strait Islander peoples throughout the data lifecycle, from the conceptualisation and creation phases of how data is defined and collected, through to how it is being analysed and used.

If resolved, PLIDA's identification of Aboriginal and Torres Strait Islander people through core datasets provides opportunities to improve and enhance research and reporting of targets and indicators specified in the National Agreement. For example, reporting on the targets related to education and employment pathways of Aboriginal and Torres Strait Islander people in Australia, a key area of focus for the National Agreement on Closing the Gap (NCVER 2023; Nicolaou et al. 2023), is currently only possible in the Census years, where the Indigenous status indicator in Census is used.

Report on Government Services

The Report on Government Services (RoGS) provides comprehensive information on the equity, efficiency and effectiveness of government services in Australia. The annual report is used by governments to inform planning and evaluation of policies, for budgeting (including to assess the resource needs and performance of government agencies) and to demonstrate government accountability. Performance across each service area is assessed using a performance indicator framework focussed on equity, effectiveness, efficiency, and outcome indicators.

The RoGS is overseen by a Steering Committee comprised of representatives from the Australian, state and territory governments. The PC is responsible for developing and producing RoGS in collaboration with representatives from relevant agencies within the Australian, state, and territory governments. Current reporting of performance indicators uses a wide range of data sources, including administrative data collections, surveys and assessment results. Linked datasets, such as PLIDA, have been identified as a notable opportunity to enhance performance reporting in RoGS¹³ (box 2.4), though any future changes to the RoGS will require the endorsement of the Steering Committee.

A potential hurdle for the success of such activities is the lag associated with data custodian approvals and output clearance. While the clearance process is important, the PC believes that there is an opportunity for PLIDA to be used for routine or regular clearance processes, provided the processes meet data privacy and security requirements.

Box 2.4 – Using PLIDA for measuring performance reporting in RoGS

Using PLIDA as a data source to enhance performance reporting in RoGS has the potential to provide three key benefits:

Enhance understanding of service use and outcomes including for specific groups

RoGS reports on service use and outcomes for specific equity groups, where data is available. Use of PLIDA or other linked datasets may provide an opportunity to improve reporting for equity groups, including Aboriginal and Torres Strait Islander people (subject to the qualifications set out above), people with disability, people from low socio-economic status and from remote/very remote locations. One example is the National Disability Insurance Scheme (NDIS) data in PLIDA, which may help with improving the identification of people with disability using mainstream government services. This data could improve reporting in RoGS on outcome areas identified in Australia's Disability Strategy 2021–2031 Outcomes Framework.

Expanding outcome indicators across different service sectors

PLIDA brings together data from multiple sources to better understand the interactions people have across different government services and their outcomes following service provision. Some examples include:

- *employment outcomes* - linkage of Vocational Education and Training (VET) completions and NDIS participants with Single Touch Payroll (STP) data
- *health service use* - linkage of NDIS participants with the Medicare Benefits Schedule (MBS) data for information on general practitioner or health specialist visits and types of treatment accessed.

Improve performance data on efficiency by linking expenditure with outcomes

Reporting on efficiency indicators measures how well government expenditure is achieving intended outcomes. Analysis of time-series data in PLIDA may enable the monitoring of changes in outcomes results in line with changes in government expenditure. A potential example is using PLIDA to examine the impact of changes in government expenditure in the VET sector on the workforce in VET-graduate

¹³ Productivity Commission 2024, Council on Federal Financial Relations review of the Report on Government Services and Performance Reporting Dashboard: RGSP Secretariat submission, www.pc.gov.au/research/supporting/rogs-performance-dashboard-review/rogs-performance-dashboard-review.pdf (accessed 26 September 2024)

Box 2.4 – Using PLIDA for measuring performance reporting in RoGS

specific occupations. Similarly, PLIDA may also provide better insights into the employment outcomes of NDIS participants.

Opportunities for future development

Currently, PLIDA linkages include a limited number of state and territory government datasets, some of which are national data collections where state and territory governments are the data owners. Future expansion of state and territory datasets may provide further opportunities for development and enhancement of performance reporting in RoGS, particularly in relation to access and use of state and territory government services, as well as potential efficiencies in data collection and report production.

3. Understanding PLIDA is essential for productive use

Already an extensive data asset, PLIDA continues to expand and deepen. To use this growing data asset effectively it is important for researchers to understand how PLIDA is structured.

Person Linkage Spine

One of the key components of PLIDA is the Spine, which is the mechanism by which one administrative dataset in PLIDA is linked to another. The Spine is what joins the records for a deidentified individual in one dataset to the data for the same deidentified individual in another. Because Australia does not have a national database that provides a unique identification number (or ‘identifier’) for each individual, PLIDA is structured around a ‘Person Linkage Spine’ that is built from key information recorded about each individual in three high quality datasets - the Medicare Consumer Directory (MCD) of Services Australia and the Department of Health and Aged Care (DoHAC), DOMINO¹⁴ from the Department of Social Services (DSS), and PIT records from the ATO. Using key variables from these core datasets (e.g. name, date of birth, gender, and address), a staged process is used to create a unique Spine identifier that links records in one dataset to those in another (Frazer 2020).¹⁵ This allows PLIDA datasets to remain separated, which helps to protect the security of the data, and the privacy of individuals recorded in it.

While approximately 48% of people in the Spine have records in all three databases, a smaller proportion of people have information from either one or two source datasets (figure 3.1). The 2023 Spine covers individuals who are Australian residents between January 2006 and June 2023.¹⁶ Those individuals who

¹⁴ DSS data asset accessible through PLIDA is referred to as DOMINO (PLIDA version).

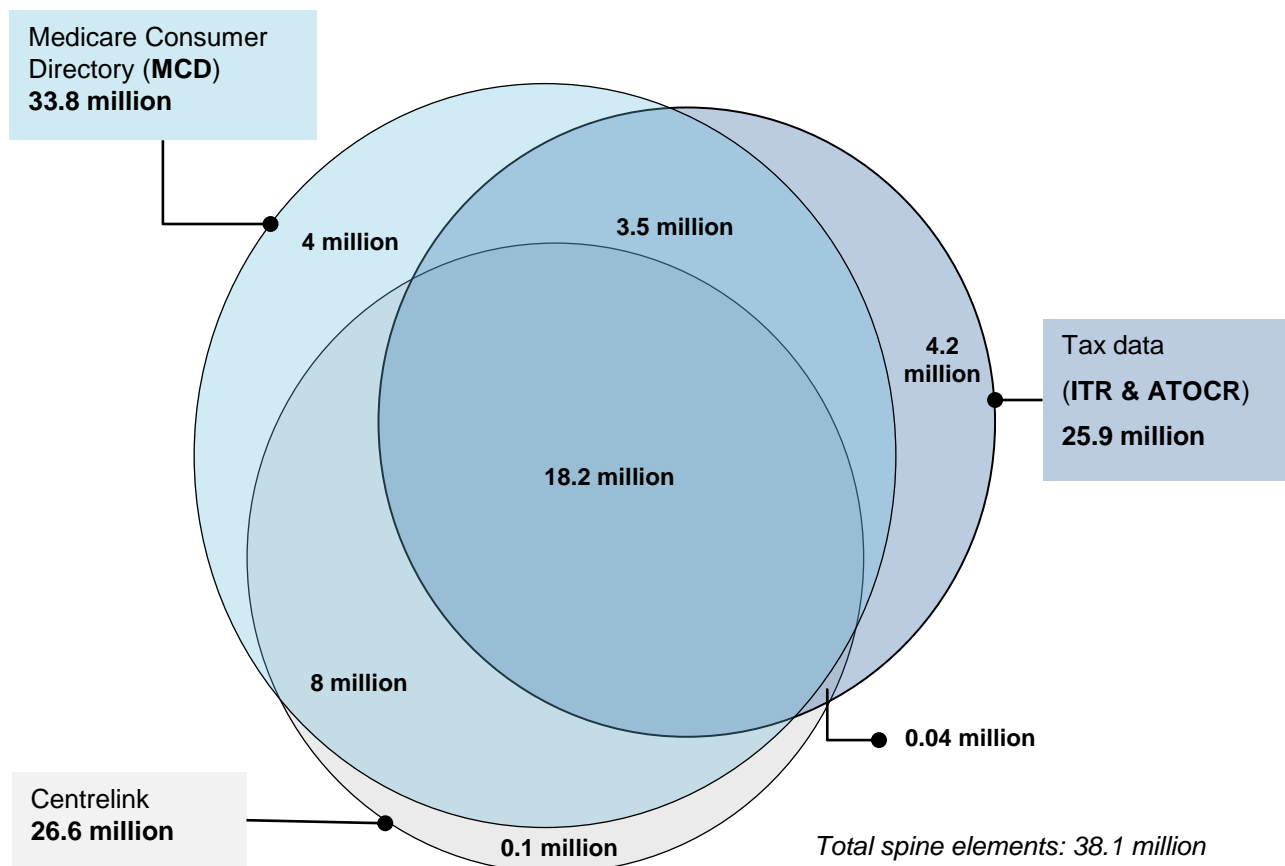
¹⁵ Detailed information about the Spine linkage process can be found in the DataLab environment. Access to this information is important to enable researchers to understand the reliability of links (Gilbert et al. 2018). Detailed information about the Spine linkage process can be found in the DataLab environment. Access to this information is important to enable researchers to understand the reliability of links (Gilbert et al. 2018).

¹⁶ The Spine is updated in December covering the population as of June of that year.

have not interacted with any of these three databases are not included in the Spine. These include, but are not limited to, recently arrived and temporary migrants, young children who are not recorded in the MCD as having received health services like vaccination, and some people aged 75 or above, who are not required to interact with the ATO and the DSS like self-funded retirees.

Figure 3.1 – The Spine is constructed from three core datasets

June 2023 Spine^a



a. The June 2023 Spine includes data from January 2006 to June 2023.

Source: ABS (2023); Wright (2021).

Since all data in the PLIDA Modular Product are deidentified, several types of unique identifiers can be used to connect individual data tables within and between datasets and modules. These include the Synthetic Agency Entity Unique ID (AEUID), Spine AEUID, and other entity IDs.

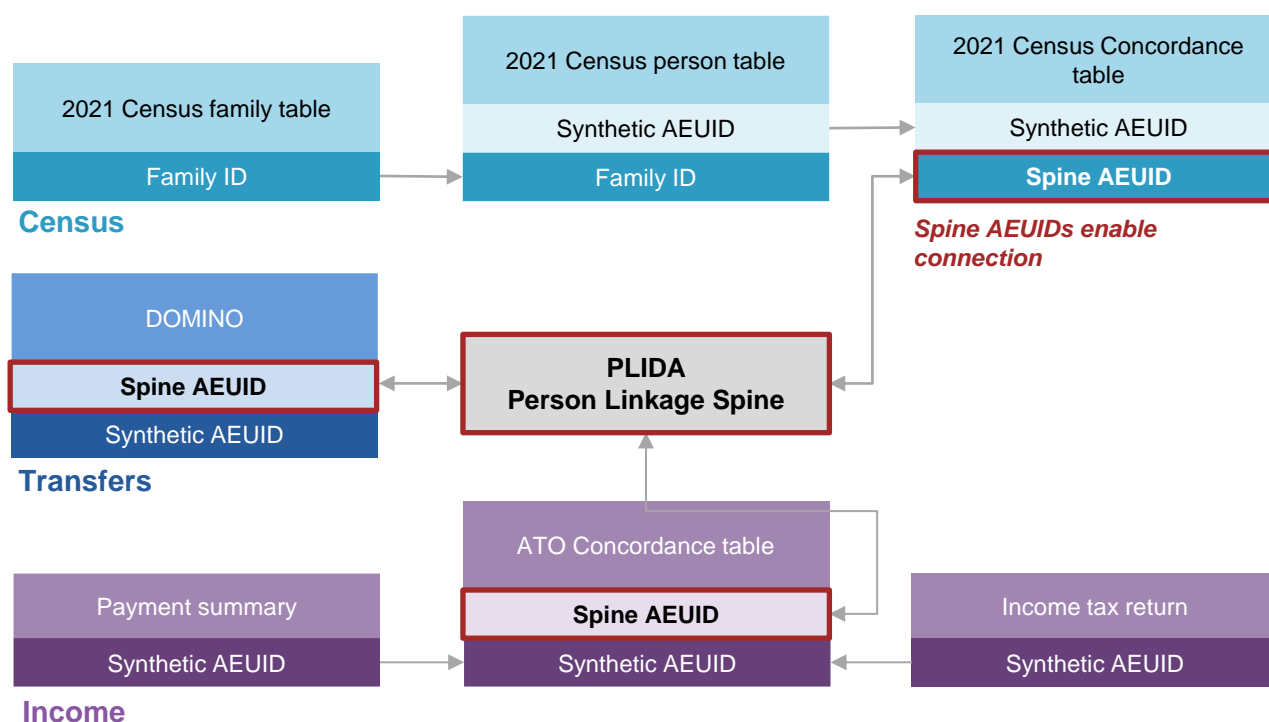
The ABS transforms unique identifiers in the original dataset into a Synthetic AEUID, enabling the datasets within PLIDA modules to be connected with one another, while preventing re-identification in the dataset outside of it. Synthetic AEUIDs are unique to an entity within a data source and are not repeated across sources.¹⁷ For example, there is no overlap between the Synthetic AEUIDs in the ATO’s module and those in Census data. Concordance tables connect each of these different Synthetic AEUIDs to corresponding Spine AEUIDs (figure 3.2).

¹⁷ Different datasets from the same data source have consistent AEUIDs, for example, a tax record with the same TFN across different datasets will have a consistent AEUID.

The Spine linkage process allocates a Spine AEUID to linked Synthetic AEUIDs within the three datasets that are used in the Spine. All datasets within the PLIDA Modular Product suite are joined through each dataset's concordance tables that consists of the Spine AEUID and a set of Synthetic AEUIDs. For example, in the PC's inequality analysis, a family ID in the Census was used to connect individual's data with family related information. For individual income data, payment summary and income tax return data tables are joined via Synthetic AEUID within the ATO dataset. Then, all information in the Census, DOMINO dataset, and ATO dataset are connected to each other via Spine AEUID which is unique to each individual (figure 3.2).

Figure 3.2 – Linking datasets within and between modules

Family incomes can be estimated using income (tax and transfer) and Census data

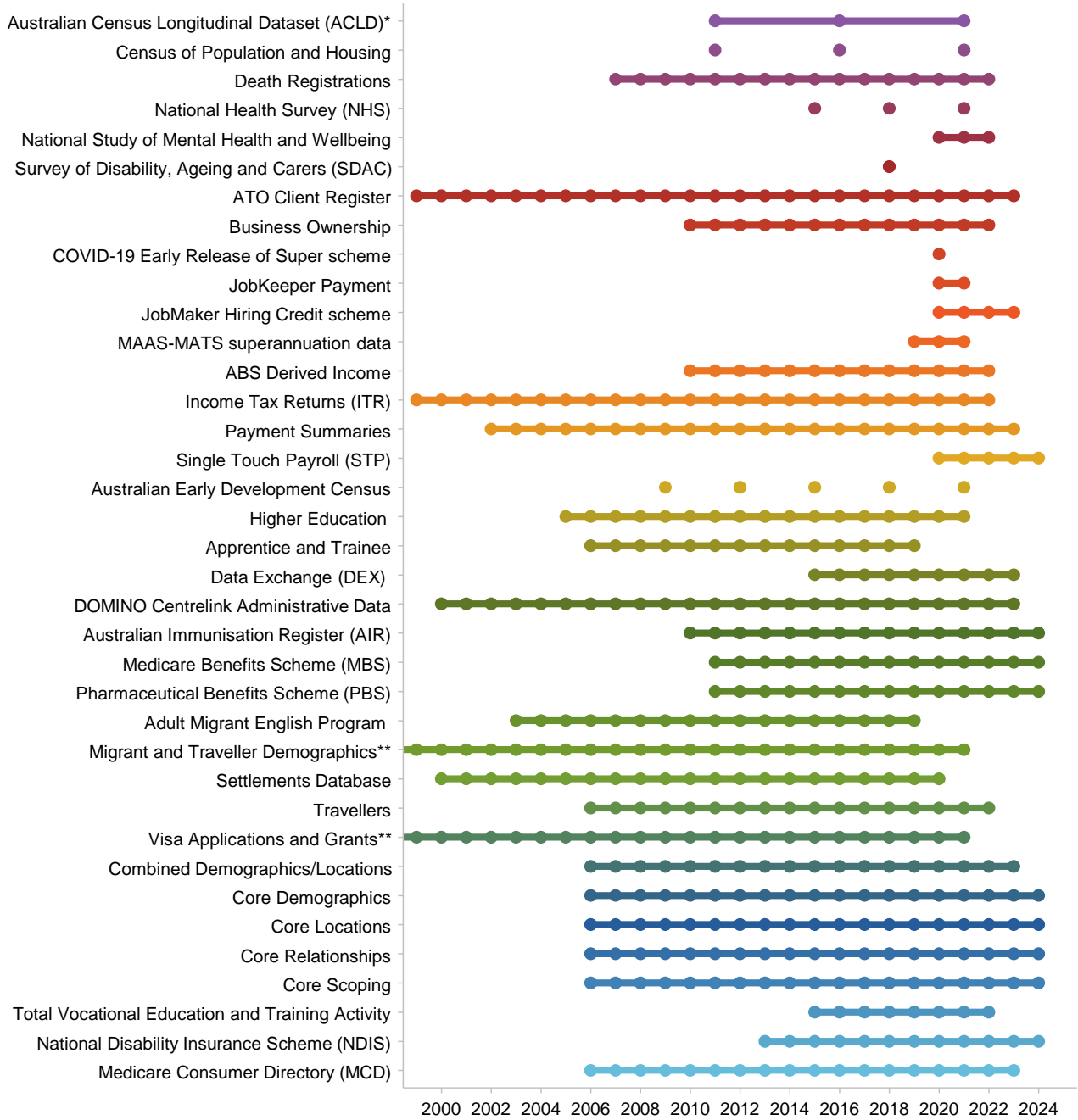


Source: Based on PC's work for the Inequality project (PC 2024a, 2024b).

PLIDA's data and time coverage

As of October 2024, PLIDA's modular product list contains 37 datasets and 270 modules. Though the majority of individual modules are longitudinal in nature, there are differences in the time coverage of each module that researchers need to be aware of. Coverage of the datasets vary, typically ranging from 1999–2024. An exception is the Migrant and Traveller Data and Visa Applications and Grants which starts from 1990 (figure 3.3).

Figure 3.3 – PLIDA modular products by datasets and by time coverage
As of September 2024



* The Australian Census Longitudinal Dataset (ACLD) is 5% longitudinal sample of Census for 2011 to 2021. ** Data is available from 1990.

Source: (ABS 2024b)

Core modules

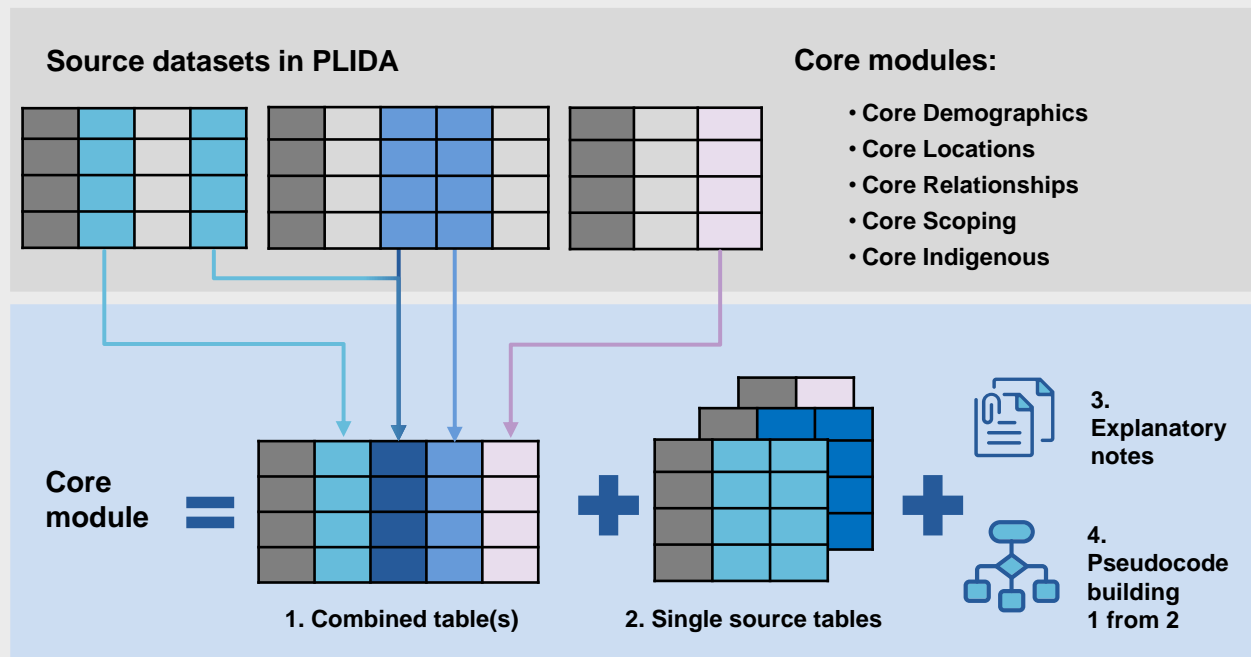
The ABS is moving to providing Core modules, along with developing additional modules (box 3.1). The aim is to assist users to identify households and families, and to more straightforwardly identify populations of interest. This would remove the need for new researchers to construct core datasets themselves and would also provide a more consistent starting point for researchers.

Box 3.1 – Core modules in PLIDA

The ABS commenced releasing the Core module datasets in July 2024, superseding the previous Combined Demographics / Locations modules. These Core modules combine information from multiple administrative datasets with the same variable (say date of birth, address) to return a single value for the individual. This is beneficial to simplify access to common variables, enables the filling of gaps in missing information in a dataset using other sources, reduces duplicative efforts in large joins, and works to ensure consistency across researchers. The Core modules also help with data minimisation, as users who previously needed to access an entire dataset for some variables could instead use the Core module values. In addition, pre-approvals by data custodians enable researchers to more readily access Core modules on generally shorter timeframes than might otherwise be the case.

The current modules include Core Demographics, Core Locations, Core Scoping, Core Relationships and Core Indigenous. Future additions may include highest educational attainment and derived income variables (see figure).

PLIDA's Core modules



The Core modules are built on the relevant information for individuals on the Spine from 1 January 2006.

- *Core Locations* module has information about an individual's location in geocoded areas and dwelling.
- *Core Demographics* module is built on the demographic information such as date of birth, gender, and country of birth.

Box 3.1 – Core modules in PLIDA

- *Core Scoping* module is designed to support users to scope PLIDA records down to an approximate population for a specific period of time and contains information about vitals (birth and death), residency/presence in Australia, and ‘activity’ across the datasets.
- *Core Relationships* module contains partnership and parent-child relationship information.
- *Core Indigenous* module includes Indigenous status indicators from some data sources along with explanatory resources to support users to navigate quality and governance considerations relating to Aboriginal and Torres Strait Islander data.

The ABS creates these modules using a rules basis that deals with conflicts or duplication. This rules approach might not be suitable for some researchers, to which the ABS offers the single source tables used for the Core module, along with their rules and pseudocode for creation. Users can create their own rules for combining the source data, which could be shared in the Code Sharing Library.

Source: ABS, personal communications, 8 November 2024.

4. Navigating some of PLIDA’s limitations

As with all datasets, PLIDA is not without its limitations. Many of these reflect limitations with the data itself, typically due to the fact the administrative data is collected for non-research purposes. This can mean the quality, coverage, or accuracy of data can vary (Amaya et al. 2020; Harron et al. 2017; Hashimoto et al. 2014). Other limitations are technical, things that make administrative data tougher to use such as computer processing requirements or dealing with complex or missing data.

Coverage limited to data relevant population

The coverage and extent of administrative datasets can often be misunderstood. While administrative datasets are much larger than traditional surveys of a limited number of people, they do not cover all people, or the coverage sometimes may be biased to certain groups of the population (Bernard et al. 2024; Curchin and Edwards 2021). That is, administrative data is not without *sparsity* concerns. While appearing large, these data cannot be automatically considered representative of the entire population, and therefore should not be considered a panacea for coverage concerns.

There may be a substantial portion of the Australian population considered ‘missing’ for certain analyses. Compared to surveys, which may aim to target a ‘representative sample’ of individuals, there are circumstances in administrative data where individuals may not be covered fully. For instance, previous research has found that up to 40% of Australian adults did not lodge a tax return, and these individuals are typically at the bottom of the personal income distribution (Atkinson and Leigh 2006). A study of the early version of PLIDA (MADIP BLE) found a similarly large proportion of adults (32%) are not in the data as a result of not having to lodge a tax return (Biddle et al. 2019). Adding Payment Summary data from employers has enabled some of these people to now be included in PLIDA.

These coverage issues might be particularly relevant in some geographic areas, like regional or remote communities. There may also be individuals who have not interacted with the relevant government services to attach them to the person linkage Spine. These include, but are not limited to, recently arrived and temporary migrants and young children (Bernard et al. 2024).

Coverage limitations from even a small proportion of the population can skew analysis. For example, were the incomes of 98% of individuals to be captured in a dataset, it would still lead to a skewed representation of the population's income distribution if the missing 2% held a significant portion of total income. Coverage issues should be considered along with other factors such as the treatment of data,¹⁸ or contextual factors like tax planning strategies involving the use of trusts, which may further impact how data is represented.

Academics have attempted to ameliorate coverage issues by combining administrative data with other data generated from targeted surveys, income tax records and transfer system data (Biddle and Marasinghe 2021). The PC undertook a similar exercise in its most recent inequality research, linking DOMINO data to ATO taxpayer data to create a more complete coverage of the population (box 2.2). When assessing data coverage challenges, it is important to ask questions of the data chosen for individual research interests, such as:

1. Who is the target population?
2. What population is represented in the data?
3. Is the target population sufficiently represented in the data?

Challenges and possible approaches

Poor coverage of a target population can skew analysis. These coverage issues can be mitigated with corroborating records across various datasets. For example, the ABS has created a close possible match to the official ABS population estimate (25.7 million in 2021) using PLIDA. It has done so by combining migration datasets, with births and deaths records to generate an indicative population estimate. The ABS's assembly team is also working towards using experience from the Core modules construction to fill coverage and data gaps in individual streams. The ABS is currently investigating whether births, deaths and migration data should be part of the Person Linkage Spine.

The PC's inequality work addressed income coverage of individuals who do not lodge tax returns by connecting DOMINO to ATO data (box 2.2). Combining data in this way can help reduce coverage issues, though this can be quite a difficult exercise for an inexperienced researcher to perform. Overall coverage issues, should be considered and flagged when conducting research.

Identifying households and families

PLIDA, as the name suggests, is a *person-level* dataset, focused on individuals. This can introduce some complications when looking to group these individuals into households or families. Household details are collected by the Census every five years. Data restrictions, and irregularities in data availability, currently stands in the way of a clear method for creating households across time. This can reduce researchers' ability to estimate variables like household equivalised income measures for inequality and poverty indicators. In an attempt to do so, the PC's most recent inequality study assumed that household size and composition

¹⁸ For example, to maintain personal privacy, individuals' personal incomes that are above (or below) certain amounts are top- (or bottom-) coded to certain thresholds. Details can be found in the PMP Explanatory Notes in the DataLab.

remained the same for a year before and after the Census year (PC 2024a). The ABS is planning to incorporate household and family composition in the PLIDA Core modules to help address some of these challenges (section 5).

Challenges and possible approaches

PLIDA is an individual centred dataset which means it often has irregular or unclear information about an individual's household or family situation. Existing data is only updated every five years with Census details. This can make specific family or household-related analysis more difficult or strongly caveated.

The PC's most recent inequality study used household data from the Census to calculate equivalised incomes. To tackle data infrequency, household location was assumed to remain the same for a year before and after the Census year due to limitations on regular data on households (PC 2024a).

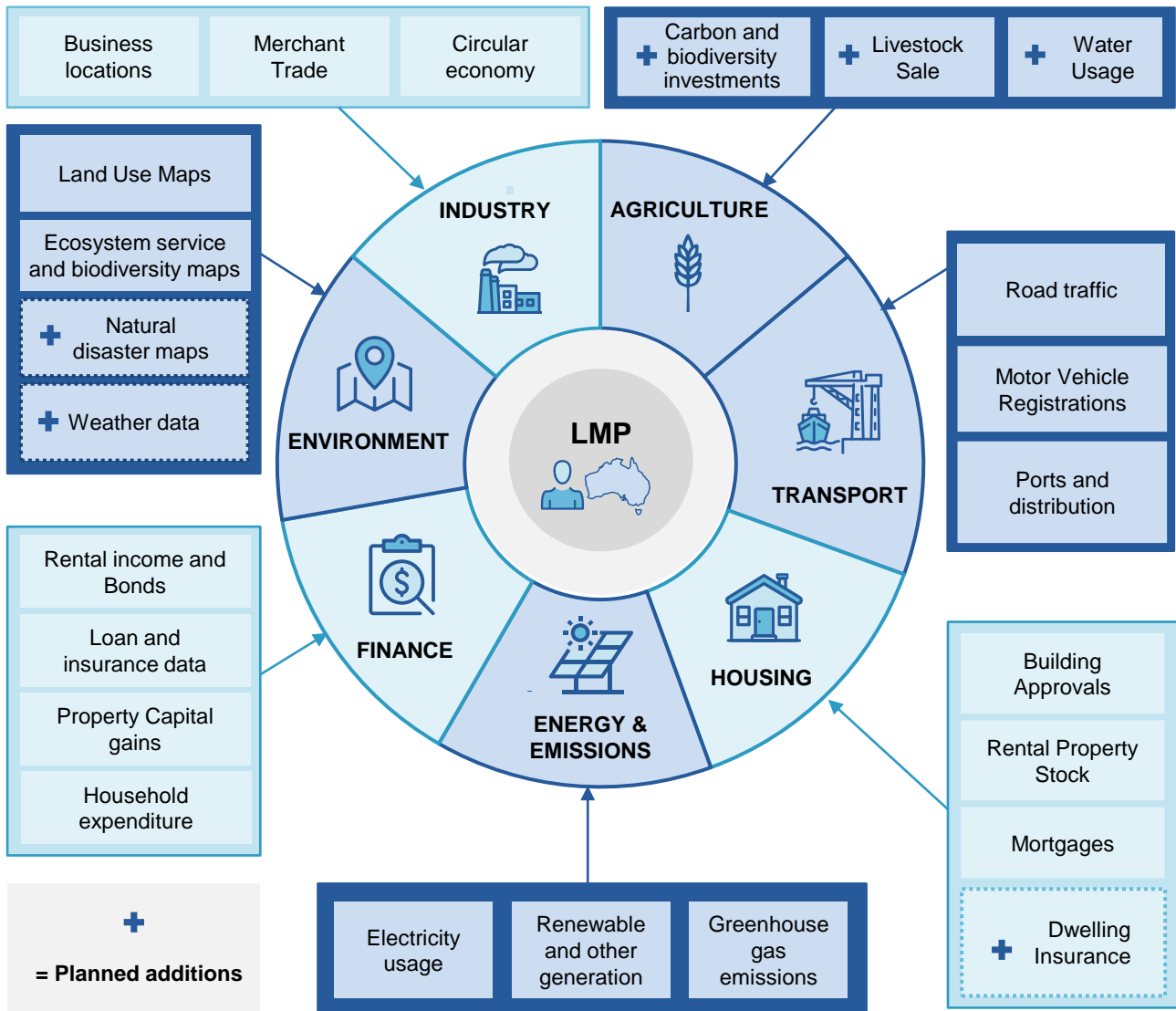
Location data

Location data can be difficult where there is not a regular and clear source of an individual's location. Data can be particularly messy where there are multiple data sources, and individuals have moved multiple times between reporting instances. Reported addresses in some datasets may be significantly out of date, like was seen in the artificially high degree of interstate migration reported as a result of Medicare address updates during the COVID-19 vaccination rollout (Centre for Population 2022, p. 54). A significant rate of spatial mismatch between Census and PLIDA for small geographical units was identified, which is prevalent among international students, recently arrived and working holiday migrants, First Nations Australians and renters (Bernard et al. 2024).

The linkable nature of administrative data enables the cross-referencing of location information from different sources, which can be both a blessing (for verifying) and a curse (in the event of conflicting information). Data like the Core Locations module can help to mitigate these problems but involve rules-based assumptions. The Spine has longitudinal address information, where each individual has historical address information associated with them which can be used to link data across time.

The ABS is proposing to harness different sources to create high quality location modular products which will enable linkage to other useful geographic data sources like industry, environment, transport, housing and energy (figure 4.1).

Figure 4.1 – Proposed Location Modular Product^a



Linkage between modules dependent on data source geography

Possible levels of Geographies:

- Address Register ID (ARID)
- Australian Statistical Geography Standard (ASGS): Meshblock to SA4
- Cadastral IDs
- Census Enumeration Geography
- Correspondences between Geographies

a. Diagram is representative of a potential future state that has been identified by the ABS. Building the LMP is iterative, as location-based datasets are identified, approved and funding secured they will be added to the LMP to be made available to approved projects and researchers.

Source: ABS, personal communications, 11 November 2024.

Challenges and possible approaches

Users can leverage data like the Core Locations, which consolidates relevant location data from a number of data assets. The Spine has longitudinal address information which can also be linked. However, even with these data there may be issues, such as where an individual has lived in multiple locations across a calendar or financial year. In such cases, researchers can use rules like the ‘most lived’ residence across the observational period or where household location data is infrequent, addresses could be simply assumed to remain the same over a certain period. For the PC’s inequality research, addresses were assumed to remain the same for a year before and after the Census year (PC 2024a).

Computational hurdles

Computational hurdles are commonplace when analysing large datasets. Prospective researchers should be aware of the computational requirements for certain activities, along with the time and cost required for these processes. While processing speeds improve notably each year, iterative, trial and error approaches to analysing large administrative data assets can still prove time consuming over the course of a given research project. Where possible, it can be useful to operate and test on a small sample of the dataset before running more lengthy processes out-of-hours.

While administrative data is generally large, certain datasets can be particularly unwieldy. For example, there are over a billion rows in the Aged Care Pension dataset for a single payment type. When looking to aggregate these payments, it very quickly becomes upwards of billions of rows. The memory in the standard machine provided by the ABS’s DataLab may not be able to read in certain datasets by default and may need to be upgraded. In cases like this, it is particularly important to bed down the required variables of a dataset early. Adding variables to a previously constructed dataset can be a time-intensive exercise if it is needed to be repeatedly re-done over the course of a project.

Challenges and possible approaches

Administrative data involves significant data joining, aggregating or wrangling, which can be computationally demanding and add to project timings. Datasets may also be too large for computer memory capacity, increasing as they are joined. Researchers should be aware that these computational demands commonly add to project timelines and cost, relative to smaller data exercises.

These concerns can be partially mitigated by using faster reading and writing file formats, such as the R package `{arrow}` (Richardson et al. 2023). Out of memory processes (for when the data is greater than the computer memory) can be used in R with the `{duckdb}` package (Mühleisen and Raasveldt 2024). It is also recommended that a researcher explore using parallel processing, or backend functions which greatly increases the speed of processing, in R this can be done with packages `{dplyr}`, `{foreach}`, or `{parallel}` (Microsoft and Weston 2022; R Core Team 2023; Wickham et al. 2023).

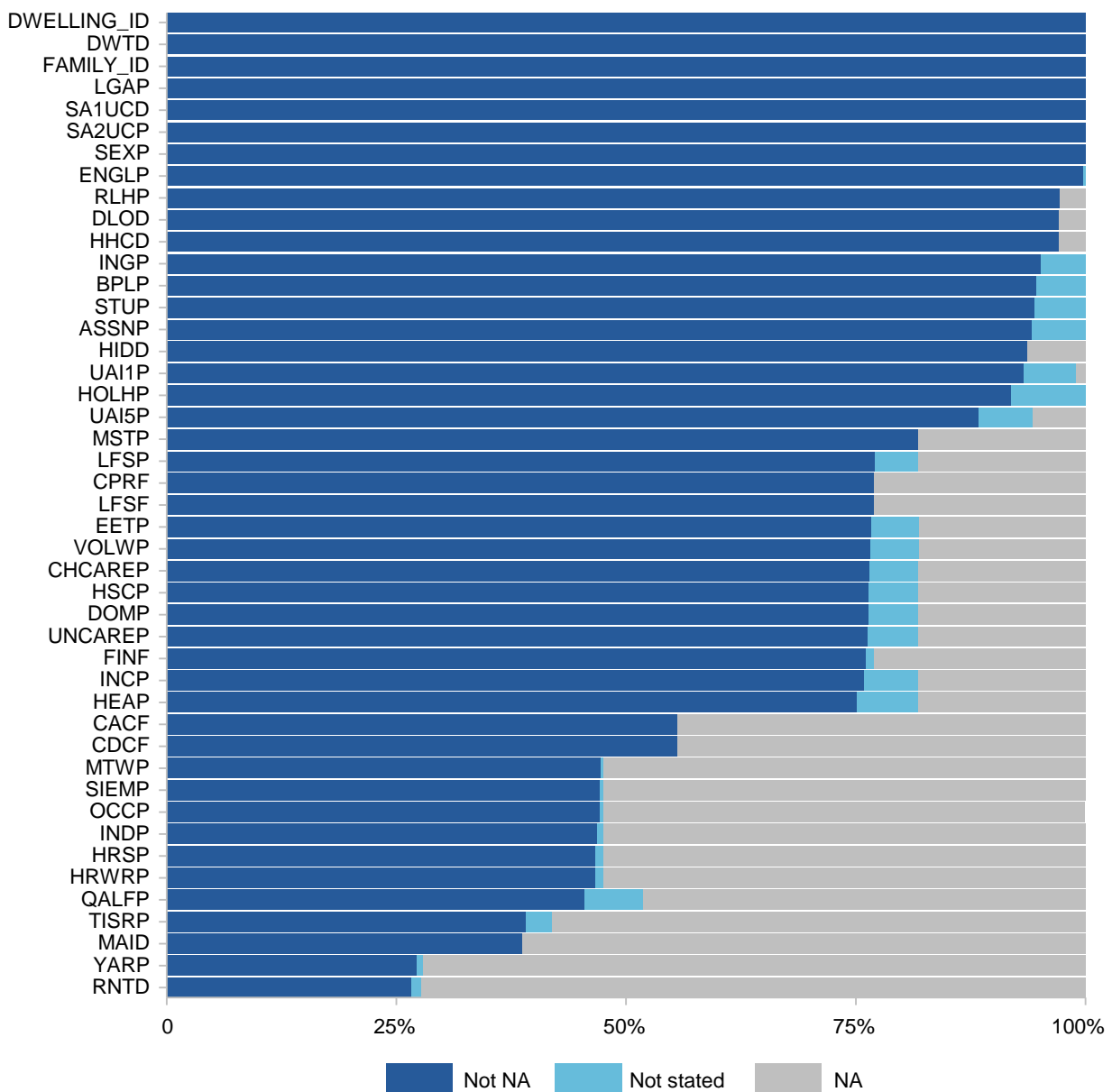
Dealing with missing data

While PLIDA covers a larger population than survey data, it can still contain missing values and there can be errors in values that are included. Missing data may be because the variable is not relevant to a unit, or it

may be missing across one dataset and available with another. For example, PLIDA's linkage with BLADE can provide a comprehensive view of the Australian economy but also introduce issues of missing data which needs to be addressed properly (McMillan and Burns 2021). In PLIDA, while the proportion of missing values is not significant (about 1%), missing values are not randomly distributed in the population. Recently arrived and temporary migrants are significantly more likely to have missing records (Bernard et al. 2024).

It can be unfair to present a number of these columns as 'missing' (figure 4.2). Data can be considered 'missing' when they are not necessarily 'missing.' For example, if an individual is not eligible for transfer payments you would not consider that value to be 'missing', simply not applicable to the individual. A similar example is that a substantial number of variables related to Higher Education Contribution Scheme (HECS) are not linkable to the Spine ID when a student does not receive HECS support.

Figure 4.2 – The number of NA values differ across variables^a



a. NA refers to a cell where no value is available, excluding those that are specified as 'not stated'.

ASSNP – Core activity need for assistance; **BPLP** – Country of birth of person; **CACF** – Count of all children in family;

CDCF – Count of dependent children in family; **CHCAREP** – Unpaid child care; **CPRF** – Count of persons in family; **DLOD** – Dwelling location; **DOMP** - Unpaid domestic work: number of hours; **DWELLING_ID** – Dwelling id variable; **DWTD** – Dwelling type; **EETP** – Engagement in employment education and training; **ENGLP** – Proficiency in spoken English; **FAMILY_ID** – Family id variable; **FINF** – Total family income (weekly); **HEAP** – Level of highest educational attainment; **HHCD** – Household composition; **HIDD** – Household income derivation indicator; **HOLHP** – Whether has other long-term health condition(s); **HRSP** – Hours worked; **HRWRP** – Hours worked (ranges); **HSCP** – Highest year of school completed; **INCP** – Total personal income (weekly); **INDP** – Industry of employment; **INGP** – Indigenous status; **LFSF** – Labour force status of parents/partners in families; **LFSP** – Labour force status; **LGAP** – Local government areas (2021 boundaries) (ur); **MAID** – Mortgage affordability indicator; **MSTP** – Registered marital status; **MTWP** – Method of travel to work; **OCCP** – Occupation code; **QALFP** – Non-school qualification: field of study; **RLHP** – Relationship in household; **RNTD** – Rent (weekly) dollar values; **SA1UCD** – SA1 (en); **SA2UCP** – SA2 (ur); **SEXP** – Sex; **SIEMP** – Status in employment; **STUP** – Full-time/part-time student status; **TISRP** – Number of children ever born (ranges); **UAI1P** – Usual address one year ago indicator; **UAI5P** – Usual address five years ago indicator; **UNCAREP** – Unpaid assistance to a person with a disability health condition or due to old age; **VOLWP** – Voluntary work for an organisation or group; **YARP** – Year of arrival in Australia.

Source: ABS DataLab (Census).

There is a significant body of work in imputing missing information from administrative datasets (Biddle and Marasinghe 2021; Chien et al. 2020; McMillan and Burns 2021; Suresh et al. 2019). These approaches can face issues where it is difficult to prove that data is missing at random. Using imputed data can paint a more complete picture of the data but users must consider the trade-off between completeness, simplicity and rigour. The creation of weighting factors can be another solution to obtain representative results (Bernard et al. 2024).

Challenges and possible approaches

Missing data across variables can present an issue where the value is genuinely missing, as opposed to simply not being applicable to the individual record. Understanding the implications of a missing value is important along with the ability to create filtering mechanisms for outliers or erroneous data. Missingness considerations will vary between variables, so users should consult the relevant documentation of key variables. If necessary, imputation methods can be used to fill missing values in large datasets or introducing weights might improve representation (Bernard et al. 2024; Chien et al. 2020; McMillan and Burns 2021; Suresh et al. 2019).

Program changes and/or requirements for data collection

Administrative data can be susceptible to program changes, merging or cessation of programs, that may affect the quality and availability of data. For example, the eligibility criteria or the service delivery model of a program may change over time, resulting in different interpretations across years. To deal with these data inconsistencies, it is important to document and understand program changes that may affect the data quality and interpretation of variables. Data dictionaries should include information about program changes or significant ‘admin data shocks’ that may influence the interpretation of the data.

For example, Higher Education data needs to be used with caution as a result of the move from Higher Education Information Management System (HEIMS) to Tertiary Collection of Student Information (TCSI). The new TCSI does not use a reporting method that directly links a course completion record with a course enrolment record. As a result, there might be insufficient data to accurately report on course completion (Australian Government 2019).

More generally, understanding the nuances of program administration can be important. For example, it may be important to know whether the childcare subsidy recorded against a given child and week is that which was

paid in the week (on the basis of eligibility as assessed at the time) or that which the customer was ultimately entitled (which may be different on the basis of reconciled income after the end of the income year).

Challenges and possible approaches

Structural breaks and or caveats in the data can impact the robustness of analysis based on that data. Researchers should be familiar with explanatory notes for each dataset in the PLIDA Modular Product (PMP) Portal. Literature reviews of existing research using the asset can also be referenced to understand common treatments of data. Clarification may need to be sought with the ABS or relevant data custodian, any clarifications should be incorporated into dataset documentation in the future.

Custodian clearance processes

Use of PLIDA data involves briefing individual data custodians of project proposals involving their datasets. While this is essential to ensure that data is used correctly, it can add time and uncertainty to projects. It is important that prospective PLIDA users understand these factors and build in sufficient time buffers into project timelines. As the scale of integrated data assets like PLIDA continue to grow, along with the number of researchers seeking to use them, these delays may lengthen as more requests are made of data custodians and prospective PLIDA users request more data for each project. The PC is aware of this challenge for their proposed use of PLIDA in ongoing reporting. Thought will need to be given to how the growth in the number of PLIDA users can be accommodated so as not to drive notable growth in clearance times. The PC understands that the ABS is developing materials that give researchers a better idea of project access and timelines and is working with custodians to streamline approval processes within current legislative requirements.

Challenges and possible approaches

Data custodians are required to respond to a data request in a reasonable time period. It is important to be conscious of these timeframes early in project planning, and to establish active lines of communication with the custodian and ABS data contact to ensure that information provided to custodians and clearance teams is correct and clearly communicated.

5. Ongoing development continues to support better access and greater use

Researchers face some costs when using PLIDA. These include financial costs, the time and capability development costs of analysing PLIDA, and the time costs of having research outputs approved for release. The largely fixed cost nature of these costs can be expected to weigh disproportionately on smaller organisations. Recent directions to reduce these costs are expected to benefit public policy researchers to that degree, potentially improving the quality of public policy conversations in the process.

Sharing opportunities between PLIDA researchers

A notable development has been the release of the PMP Portal in DataLab. The PMP Portal acts as an information repository that includes explanatory notes, a data dictionary, and a statistical wiki. It also includes a new *Code Sharing Library* to enable researchers to share their code or processes. Collectively, these stand to build a community of practice around PLIDA over time, helping researchers to reduce the time costs of their projects. The PMP Portal's *Code Sharing Library* is still in its infancy but is expected to be expanded as researchers begin to submit code. As an initial step, the ABS has provided sets of code, developed by the Australian Institute of Family Studies (AIFS), for measures of family relationship and government-provided income from the DOMINO dataset. The PC also understands that an additional set of codes for processing STP data has been submitted by Treasury researchers for inclusion in the Code Sharing Library which other data users have expressed an interest in using for their research. This is a clear indication of how important and significant the code sharing in data processing and research are. Code sharing promises to reduce the time cost of data wrangling and cleaning of administrative data, but also ensures that research and data techniques are transparent and reproducible.

Training resources for new users

Using PLIDA requires operating in the DataLab environment. In order to be able to apply for DataLab access a prospective user must have 'three years of either quantitative research experience ... or a referral from an experienced researcher working on the same project.' As part of the approval process for accessing DataLab, the ABS provides DataLab training which helps prospective users understand their obligations for privacy and protecting cleared research outputs. This training does not extend to how to use PLIDA itself, however. This absence of basic PLIDA-specific training differs from that made available to prospective HILDA users, which covers topics like basic infrastructure, how to use its modules, and how to extract data from those modules.

It is natural to anticipate that as demand for PLIDA use grows, opportunities and offerings for training will emerge, either from the group of registered users, or from the ABS itself. Currently, there are some discrete, topic specific training courses provided by expert researchers, such as those supported by the DoHAC. While these are not publicly available, the PC understands that there is growing interest to expand these types of training offerings. In the meantime, new PLIDA users might helpfully collaborate with existing PLIDA researchers as a way of developing this capability.

Research outputs and one-off linkages

The ability for researchers to let other PLIDA users know of their current research questions, whether they are using a one-off linkage for that research,¹⁹ or sharing their findings, can build conversations around individual research questions, motivate and inspire new ones, and save time on those subsequently pursued (PC 2017, p. 278). The ABS maintains a non-exhaustive list of *PLIDA/MADIP Research Projects*, that provides

¹⁹ In addition to existing PLIDA modules, researchers can request the ABS and data custodians to undertake discrete, one-off linkage activities for specific projects. For example, a one-off linkage was created for lung cancer research by linking census records with the NSW Cancer Registry with PLIDA (Banham et al. 2024). The PC leveraged a linkage for SIH and CCS data from the Department of Education, Skills and Employment within PLIDA for the ECEC inquiry (PC 2024a).

information about the various projects undertaken by academia and agencies, is a useful step in this direction.²⁰

Nevertheless, of the 356 projects listed on this register, only 8 projects were found to have public links to the outputs or paper associated with them.²¹ As the community of PLIDA users grows in Australia it is natural to anticipate that more systems of output sharing might emerge. In the meantime, new and existing PLIDA researchers might benefit by promoting this visibility themselves. Researchers may promote visibility either by making a concerted effort to share their research with communities of practice, requesting that other researchers and organisations be designated as discussants on their PLIDA research projects, or collaborating on research projects more directly.

DataLab's conditions of use require users to share an upcoming publication two weeks prior to its release with the ABS. This appears to be a natural mechanism to capture and collate publicly released outputs, appending them to the already listed projects. These project outputs could transition to an archived list of published projects, to ensure that historical projects and their outputs remain easily discoverable. For a user commencing research, this is an avenue to understand existing research. For a publishing user, this is a way to provide access to a wide range of interested researchers, and the public.

Standardised datasets have started to become available

Standardised datasets within PLIDA can help reduce start-up costs for researchers. The PC saw the benefits of this during its 2024 inequality research, where PC researchers leveraged a core dataset created by a team member's earlier effort (PC 2024a). In addition to avoiding duplication of effort, this approach provided for consistency in the data sources being drawn upon in the PC's inequality research.

Developing standardised datasets for popular dataset combinations could help new or existing researchers save time and provide a consistent starting point for their research. The ABS's provision of Core modules for PLIDA presents a preliminary step in this direction, enabling researchers to be satisfied by the consistency of data being used for different research projects, reducing the time otherwise required for research projects, and providing the confidence that the datasets have been joined and constructed in an appropriate way.

More state and territory data is becoming available in PLIDA

Given that states and territories are the principal provider of the many of the government services, recent moves by some states and territories to make these data available to PLIDA is a welcome development for public policy research in Australia. A notable example is the development of a dataset for the Life Course Data Initiative (LCDI), which aims to develop datasets that enable insights on the life course of individuals. This could enable research to be undertaken across a range of areas, including research into the nature of entrenched disadvantage. Under LCDI, selected South Australia and Australian Capital Territory service data is to be linked via PLIDA's Spine Linkage and are expected to be available for use in March 2025 (table 5.1).²²

²⁰ [PLIDA/MADIP Research Projects | Australian Bureau of Statistics \(abs.gov.au\)](#)

²¹ The project register may also use umbrella projects. For example, the PC's DataLab projects fall under an umbrella topic of research and reporting, not by individual project names.

²² More details about LCDI can be found at [Life Course Data Initiative | Australian Bureau of Statistics](#).

Currently, PLIDA has deaths registry data for all jurisdictions. The ABS has confirmed that linking the births registry, hospitals, and justice and corrections datasets are under development at the time of publication of this paper. There are some state and territory datasets that are linked with PLIDA for specific projects and would not be readily available for broader use due to data supply arrangements.

Table 5.1 – Existing and planned linkages of state and territory data

Datasets	NSW	Vic	Qld	SA	WA	NT	ACT	Tas
Deaths Registry	✓	✓	✓	✓	✓	✓	✓	✓
Births Registry	✓	✓	✓	✓	✓	✓	✓	✓
Cancer Registry	✓	✓	✓	✓	✓	✓	✓	✓
Justice and correction	✓	✓	✓	✓	✓	✓	✓	✓
Hospitals	✓	✓	✓	✓	-	✓	✓	✓
Health	-	-	-	✓	-	-	✓	-
Education	-	-	-	✓	-	-	✓	-
Housing	-	-	-	✓	-	-	-	-

✓ Linked ✓ Under development.

Source: ABS, personal communications, 11 November 2024.

Appendices

Appendix A – PLIDA data items

Table A.1 – PLIDA data custodians and datasets^a

Data custodian	Dataset acronym	Dataset name	Dataset description	Reference period	Update frequency
ABS	ACLD	Australian Census Longitudinal Dataset	A 5% sample of longitudinally linked 2011 and 2016 Census records.	2011 to 2021	Follows ABS publication
ABS	CENSUS	Census of Population and Housing	Information from the ABS Census of Population and Housing, including person, family, and dwelling information.	2011, 2016, 2021	Follows ABS publication
ABS	DEATHS	Death Registrations	Information on persons who have died in Australia, including cause of death and demographics.	2007 to 2022	Follows ABS publication
ABS	NHS	National Health Survey	Information about the health and wellbeing of people in Australia, such as medical conditions, health and lifestyle risk-factors, mental health, and use of health services.	2014-2015, 2017-2018, 2020-2021	Follows ABS publication
ABS	NSMHW	National Study of Mental Health and Wellbeing	Information on the mental health and wellbeing of people in Australia, including prevalence of selected mental disorders, level of impairment, health service usage, suicidality and self-harm behaviours, and demographic and socio-economic characteristics.	2020 to 2022	Follows ABS publication
ABS	SDAC	Survey of Disability, Ageing and Carers	Information about people with a disability, people aged 65 and over, and their carers.	2018	Follows ABS publication
ATO	ATO_CR	ATO Client Register	Demographic information about persons who have registered for an Australian Tax File Number with the ATO.	1999 to 2023	Annual
ATO	BUSOWN	Business Ownership	Concordances to enable MADIP data about sole traders and partnership beneficiaries to be linked to BLADE data about the businesses they operate.	2009-10 to 2021-22	Annual
ATO	ERS	COVID-19 Early Release of Super scheme	Information about individuals' requests of early release of their superannuation under the COVID-19 Early Release to Super scheme.	2020	Ceased
ATO	JK	JobKeeper Payment	Information about payments to eligible persons under the JobKeeper Payment scheme.	2020 to 2021	Ceased
ATO	JM	JobMaker Hiring Credit scheme	Information about payments made for eligible employees under the JobMaker Hiring Credit Scheme.	2020 to 2023	Ceased

Data custodian	Dataset acronym	Dataset name	Dataset description	Reference period	Update frequency
ATO	MAASMATS	MAAS-MATS superannuation data	Information about superannuation account holders and their superannuation accounts.	2018-19 to 2020-21	Annual
ATO	PIT_IE	ABS Derived Income	Information on income derived by the ABS from ITR and PS data to align with ABS income definitions.	2010-11 to 2021-22	Annual
ATO	PIT_ITR	Income Tax Return	Information from individual income tax returns submitted to the ATO, including income, losses, deductions, expenses, offsets, tax withheld and debts.	1999-00 to 2021-22	Annual
ATO	PIT_PS	Payment Summaries	Information from payment summaries provided by employers to their employees and submitted to the ATO, including payments and the amounts withheld from those payments.	2001-02 to 2022-23	Annual
ATO	STP	Single Touch Payroll	Information about individuals' payroll data from their employers.	2020 to current	Monthly
DE	AEDC	Australian Early Development Census	Information from the AEDC, a nationwide collection of early childhood development at the time children commences their first year of full-time school.	2009, 2012, 2015, 2018, 2021	3-yearly
DE	HE	Higher Education	Information about enrolments, completions, courses, and loan amounts for students studying at Australian higher education institutions.	2005 to 2021	Annual
DEWR	A&T	Apprentice and Trainee	Information about Australian Apprentices and their employers.	2006 to 2019	Not scheduled
DSS	DEX	Data Exchange	Information about clients of select programs funded by government grants, reported through the DSS Data Exchange.	2015 to 2023	Annual
DSS	DOMINO	DOMINO Centrelink Administrative Data	Information about people who have interacted with Centrelink, including demographics, eligibility for Centrelink benefits, and actual benefits received.	2000 to 2023	Quarterly
DoHAC	AIR	Australian Immunisation Register	Information about vaccinations provided to people, as recorded in the Australian Immunisation Register.	2010 to 2024	Weekly
DoHAC	MBS	Medicare Benefits Schedule	Information about claims processed for subsidised health services provided to people under the Medicare Benefits Schedule.	2011 to 2024	Quarterly
DoHAC	PBS	Pharmaceutical Benefits Scheme	Information about medicines provided to people under the Pharmaceutical Benefits Scheme.	2011 to 2024	Quarterly
HA	AMEP	Adult Migrant English Program	Information about participants in the Adult Migrant English Program, a free service to help eligible migrants improve their English language skills.	2003 to 2019	Not scheduled
HA	MT_DEMOGS	Migrant and Traveller Demographics	Demographic and location information about persons who have migrated or travelled to or from Australia	1990 to 2021	Annual
HA	SDB	Settlements Database	Information about permanent migrants who have settled in Australia.	2000 to 2020	Annual
HA	TRAVELLERS	Travellers	Information about people who have travelled in or out of Australia.	2006 to 2022	Annual
HA	VISA	Visa Applications and Grants	Information about people who have applied for visas to migrate to Australia.	1990 to 2021	Annual

Data custodian	Dataset acronym	Dataset name	Dataset description	Reference period	Update frequency
Multiple	COMBINED	Combined Demographics / Combined Locations	Demographic and location information combined from multiple sources by the ABS for ease of use by researchers.	2006 to 2023	Annual
Multiple	CORE	Core Demographics	Demographic information combined from multiple sources by the ABS.	2006 to latest	Annual
Multiple	CORE	Core Locations	Location information combined from multiple sources by the ABS.	2006 to latest	Annual
Multiple	CORE	Core Relationships	Relationships information combined from multiple sources by the ABS.	2006 to latest	Annual
Multiple	CORE	Core Scoping	Vitals, residence, and activity information combined from multiple sources by the ABS which can be used to scope the PLIDA population to a point in time.	2006 to latest	Annual
NCVER	TVA	Total VET Activity	Information about students undertaking Vocational Education and Training, collected by Registered Training Organisations.	2015 to 2022	Annual
NDIA	NDIS	National Disability Insurance Scheme	Information about participants and providers in the National Disability Insurance Scheme.	2013 to 2024	6-monthly
SA	MCD	Medicare Consumer Directory	Demographic information about persons who have registered for a Medicare number with Services Australia.	2006 to 2023	Annual

a. The full data item list for PLIDA is on the ABS website: [Microdata: Person Level Integrated Data Asset \(PLIDA\) | Australian Bureau of Statistics \(abs.gov.au\)](https://www.abs.gov.au/microdata/person-level-integrated-data-asset-plida)

Source: (ABS 2024b)

Appendix B – Who has produced what?

The ABS's PLIDA page provides a list of projects by various Australian and state and territory governments, and academic and private institutions. This table presents the list of research outputs that are cited in this paper, linking them to the institution as in the Research projects register for PLIDA.

Table B.1 – Who has produced what using PLIDA^a

Title	Citation	Institution	Datasets
Measuring Labour Quality in (Closer to) Real Time Using Emerging Microdata Sources	Hambur et al. 2024	RBA	CENSUS, DOMINO, PIT, STP
Measuring vulnerability and disadvantage in early childhood data collections: Phase two	Pham et al. 2024	DESE	CENSUS, AEDC, DEX, DOMINO, MBS, NHS, PIT, PBS, DEATHS, NQS*, CCMS*
Linking quality and child development early childhood education and care: Technical report	Rankin et al. 2024	DESE	PLIDA (MADIP), E4Kids longitudinal dataset* (Effective Early Educational Experiences data from QLD and VIC)
Access to high-quality early care and education: Analysis of Australia's national integrated data	Tang et al. 2024	DESE	CENSUS, AEDC, PIT, DEX
Australia's health 2024: data insights	AIHW 2024	AIHW	PLIDA
Social determinants of health among culturally and linguistically diverse people in Australia	AIHW 2024	AIHW	CENSUS, NHS, MBS, PBS, DEATHS
COVID-19 reporting	DoHAC 2024	DoHAC	AIR, CENSUS
A snapshot of inequality in Australia	PC 2024	PC	CENSUS, DOMINO, PIT, COMBINED
Rising inequality? A stocktake of the evidence	PC 2024	PC	CENSUS, DOMINO, PIT, COMBINED
A path to universal early childhood education and care	PC 2024	PC	PLIDA: SIH MURF is linked with administrative CCS data*
Demographic, health and socioeconomic characteristics related to lung cancer diagnosis: a population analysis in New South Wales, Australia	Banham et al. 2024	n/a	CENSUS, PBS, and NSW Cancer Registry*

Title	Citation	Institution	Datasets
The pitfalls and benefits of using administrative data for internal migration research: An evaluation of Australia's Person Level Integrated Data Asset (PLIDA)	Bernard et al. 2024	University of Queensland	CENSUS, Combined Location module
Labour market outcomes of skilled migrants in South Australia	Cebulla et al. 2024	Flinders University	CENSUS, PIT, PS, VISA, MT_DEMOGS
Using new data sources to understand and monitor changes in prices, wages and incomes	Fink, Hambur, and Majeed 2023	RBA	PLIDA
Parental Leave Pay and Dad and Partner Pay: Patterns of use. Analysis using the Person Level Integrated Dataset (PLIDA)	Baxter and Budinski 2023	AIFS	CENSUS, DOMINO
Measuring vulnerability and disadvantage in early childhood data collections	Gray et al. 2023	DESE	CENSUS, AEDC, DEX, DOMINO, MBS, NHS, PIT, PBS, DEATHS, NQS*, CCMS*
Health of refugees and humanitarian entrants in Australia	AIHW 2023	AIHW	CENSUS, SDB, MCD, MBS, PBS, DEATHS, COMBINED
Effectiveness of COVID-19 vaccination against COVID-19 specific and all-cause mortality in older Australians: a population-based study	Liu et al. 2023	DoHAC	CENSUS, DEATHS, AIR
Early pension withdrawal as stimulus	Hamilton et al. 2023	ANU – TTPI	AEDC, CENSUS, DOMINO, MBS, NHS, PIT, PBS, DEATHS
From VET to sustainable employment for Aboriginal and Torres Strait Islander peoples	NCVER 2023	NCVER	CENSUS, TVA
First Nations People Workforce Analysis	Nicolaou et al. 2023	Jobs and Skills Australia	CENSUS, PIT, TVA, HE
Understanding the wellbeing characteristics of ex-serving ADF members	AIHW 2022	AIHW	CENSUS, Combined Demographics, a one-off linkage with PMKeyS*
Understanding the fatal burden of COVID-19 in residential aged care homes in Australia: Using linked data to generate evidence	Welsh et al. 2022	DoHAC	CENSUS, DEATHS, AIR
The lifetime fiscal impact of the Australian permanent migration program	Varela et al. 2021	Treasury	PLIDA (MADIP): CENSUS, MBS, PBS, VISA

Title	Citation	Institution	Datasets
Socioeconomic determinants of vaccine uptake: July 2021 to January 2022	Biddle et al. 2021	DoHAC	CENSUS, AIR, DEATHS
Using Census, Social Security and Tax Data from the Multi-Agency Data Integration Project (MADIP) to Impute the Complete Australian Income Distribution	Biddle and Marasinghe 2021	ANU	PIT, ACLD, SSRI (Social Security and Related Information – predecessor of DOMINO)
Benefits of educational attainment	DESE 2020	DESE	PLIDA (MADIP): CENSUS, MBS, PBS, HE (HEIMS), PIT_ITR & PIT_PS (PIT/PAYG), DEATHS
Mental Health inquiry report	PC 2020	PC	PLIDA (MADIP)
Introducing the longitudinal MADIP and its role in understanding income dynamics in Australia	Biddle et al. 2019	ANU	PLIDA (MADIP BLE 2011)
Childcare and early childhood learning public inquiry	PC 2015	PC	SSRI (Social Security and Related Information – predecessor of DOMINO)

* Linked in PLIDA for a particular project. **n/a** Not applicable. **a.** The list contains research outputs cited in this paper and thus is not exhaustive list of research outputs that use PLIDA.

Abbreviations

ABS	Australian Bureau of Statistics
ACLD	Australian Census Longitudinal Dataset
ACT	Australian Capital Territory
AEDC	Australian Early Development Census
AEUID	Agency Entity Unique ID
AIFS	Australian Institute of Family Studies
AIHW	Australian Institute of Health and Welfare
AIR	Australian Immunisation Records
ALIFE	ATO Longitudinal Information Files
ANU	Australian National University
ANZICS	Australian and New Zealand Intensive Care Society
APP	Australian Privacy Principles
APS	Australian Public Service
ATO	Australian Taxation Office
ATOOCR	Australian Taxation Office Client Register
BLADE	Business Longitudinal Analysis Data Environment
BLE	Basic Longitudinal Extract
CCS	Childcare Subsidy
CtG	Closing the Gap
CVDL	Centre for Victorian Data Linkage
DESE	Department of Education, Skills and Employment
DoHAC	Department of Health and Aged Care
DOMINO	Data Over Multiple Individual Occurrences
DSS	Department of Social Services
ECEC	Early Childhood Education and Care
HECS	Higher Education Contribution Scheme
HEIMS	Higher Education Information Management System
HES	Household Expenditure Survey
HILDA	Household Income and Labour Dynamics in Australia
LCDI	Life Course Data Initiative
LLFS	Longitudinal Labour Force Survey
LMP	Location Modular Product

MADIP	Multi-Agency Data Integration Project
MBS	Medicare Benefits Schedule
MCD	Medicare Consumer Directory
MHWS	Mental Health Wellbeing Survey
NCVER	National Centre for Vocational Education and Research
NDIS	National Disability Insurance Scheme
NHS	National Health Survey
NSW	New South Wales
NT	Northern Territory
OECD	Organisation for Economic Cooperation and Development
PBO	Parliamentary Budget Office
PBS	Pharmaceutical Benefits Schedule
PC	Productivity Commission
PIA	Privacy Impact Assessment
PIT	Personal Income Tax
PLIDA	Person Level Integrated Data Asset
PMP	PLIDA Modular Product
QLD	Queensland
RBA	Reserve Bank of Australia
RoGS	Report on Government Services
SA	South Australia
SIH	Survey of Income and Housing
SSRI	Social Security Related Information
STP	Single Touch Payroll
TAS	Tasmania
TCSI	Tertiary Collection of Student Information
TTPI	Tax and Transfer Policy Institute
VET	Vocational Education and Training
VIC	Victoria
WA	Western Australia

References

- ABS (Australian Bureau of Statistics) 2021a, *Input and output clearance*.
- 2021b, *Migrant Data Matrices*.
- 2022, *ABS Submission to the Productivity Inquiry 2022*.
- 2023, *Person Linkage Spine 20062023 Methodology and Quality Assessment*, June, ABS DataLab.
- 2024a, *Keeping integrated data safe | Australian Bureau of Statistics*.
- 2024b, *PLIDA/MADIP Research Projects*.
- 2024c, *Privacy Impact Assessment: Expanded Health Data Linkage to the Person Level Integrated Data Asset*.
- 2024d, *Privacy Impact Assessments | Australian Bureau of Statistics*.
- 2024e, *Privacy in PLIDA | Australian Bureau of Statistics*.
- AIHW (Australian Institute of Health and Welfare) 2022, *Reporting on the health of culturally and linguistically diverse populations in Australia: An exploratory paper*, Australian Institute of Health and Welfare.
- 2023, *Health of refugees and humanitarian entrants in Australia*, Web report.
- 2024a, *Australia's health 2024: data insights*.
- 2024b, *Social determinants of health among culturally and linguistically diverse people in Australia*, Web report.
- Amaya, A, Biemer, PP and Kinyon, D 2020, 'Total Error in a Big Data World: Adapting the TSE Framework to Big Data', *Journal of Survey Statistics and Methodology*, vol. 8, no. 1, pp. 89–119.
- Andrews, D and Hansell, D 2019, 'Productivity-Enhancing Labour Reallocation in Australia', *Treasury Working Paper*, vol. 2019–06, no. 317.
- Atkinson, AB and Leigh, A 2006, 'The distribution of top incomes in Australia', *CEPR Discussion Paper*, p. 514.
- Australian Government 2019, *Prepopulating TCSI with HEIMS data*.
- Banham, D, Roder, D, Stone, E, Quayle, S, Rushton, S and O'Brien, T 2024, 'Demographic, health and socioeconomic characteristics related to lung cancer diagnosis: a population analysis in New South Wales, Australia', *Discover Social Science and Health*, vol. 4, no. 1, p. 34.
- Baxter, J and Budinski, M 2023, *Parental Leave Pay and Dad and Partner Pay: Patterns of use. Analysis using the Person Level Integrated Dataset (PLIDA)*, Australian Institute of Family Studies.
- Bernard, A, Wu, J, Wilson, T, Argent, N, Zajac, T and Kimpton, A 2024, 'The pitfalls and benefits of using administrative data for internal migration research: An evaluation of Australia's Person Level Integrated Data Asset (PLIDA)', *Demographic Research*, vol. 51, pp. 687–722.
- Biddle, N, Breunig, R, Markham, F and Wokker, C 2019, *Introducing the longitudinal MADIP and its role in understanding income dynamics in Australia*, CSRM & SRC methods paper, 1.
- Biddle, N and Marasinghe, D 2021, *Using Census, Social Security and Tax Data from the Multi-Agency Data Integration Project (MADIP) to Impute the Complete Australian Income Distribution*, Working paper 8/2021, Tax and Transfer Policy Institute.
- Biddle, N and Markham, F 2018, 'Indigenous identification change between 2011 and 2016: Evidence from the Australian Census Longitudinal dataset', *CAEPR Topical Issue*, no. 1.
- Biddle, N, Welsh, J, Butterworth, P, Edwards, B and Korda, R 2022, *Socioeconomic determinants of vaccine uptake: July 2021 to January 2022*, Commissioned study, Department of Health.
- Borland, J and Smedes, M 2024, 'A review of how economists thinking about human capital has evolved and the current state of research on human capital in Australia'.
- Bruno, A, Hambur, J and Wang, L 2024, 'Measuring Labour Quality in (Closer to) Real Time Using Emerging Microdata Sources', presented at Joint ABS-RBA Conference on Human Capital.
- Card, D, Chetty, R, Feldstein, MS and Saez, E 2010, *Expanding Access to Administrative Data for Research in the United States*, SSRN Scholarly Paper, Rochester, NY.
- Cebulla, A, Gamble, H and Tan, G 2024, *Labour market outcomes of skilled migrants in South Australia*, Australian Industrial Transformation Institute, Flinders University of South Australia, Adelaide.
- Centre for Population 2022, *2022 Population Statement*, Australian Government, Canberra.
- Chien, C-H, Welsh, AH and Moore, JD 2020, 'Synthetic Business Microdata: an Australian example', *Journal of Privacy and Confidentiality*, vol. 10, no. 2, pp. 1–21.
- Cole, S, Dhaliwal, I, Sautmann, A and Vilhuber, L 2021, *Handbook on Using Administrative Data for Research and Evidence-based Policy*, Massachusetts Institute of Technology.
- Connelly, R, Playford, CJ, Gayle, V and Dibben, C 2016, 'The role of administrative data in the big data revolution in social science research', *Social Science Research*, vol. 59, pp. 1–12.
- Curchin, K and Edwards, B 2021, 'Introduction to the Special Issue on Big Data and Social Policy in Australia', *Australian Journal of Social Issues*, vol. 56, no. 2, pp. 132–142.
- Department of the Prime Minister and Cabinet and Commonwealth of Australia 2024, *COVID-19 Response Inquiry Report*.
- DESE (Department of Education, Skills and Employment) 2020, *Apprenticeship graduate outcomes factsheets*.
- Deutscher, N 2019, *Job-to-job transitions and the wages of Australian workers*, Treasury Working Paper.
- DoHAC (Department of Health and Aged Care) 2024, *COVID-19 reporting*, Australian Government Department of Health and Aged Care.
- Frazer, B 2020, 'Person Spine linkage methodology and maintenance', *International Journal of Population Data Science*, vol. 5, no. 55.

- Gilbert, R, Lafferty, R, Hagger-Johnson, G, Harron, K, Zhang, L-C, Smith, P, Dibben, C and Goldstein, H 2018, 'GUILD: GUIDance for Information about Linking Data sets', *Journal of Public Health*, vol. 40, no. 1, pp. 191–198.
- Gray, S, Pham, C, Guo, J, Downes, M, O'Connor, E, Priest, N and Goldfeld, S 2023, *Measuring vulnerability and disadvantage in early childhood data collections*, Centre for Community Child Health, Melbourne.
- Griffiths, K, Coleman, C, Al-Yaman, F, Cunningham, J, Garvey, G, Whop, L, Pulver, LJ, Ring, I and Madden, R 2019, 'The identification of Aboriginal and Torres Strait Islander people in official statistics and other data: Critical issues of international significance', *Statistical Journal of the IAOS*, vol. 35, no. 1, pp. 91–106.
- Gruen, D 2024, 'Data linkage and integration to improve the evidence base for public policy: lessons from Australia'.
- and Goldbloom, A 2008, 'The smarter use of data'.
- Hambur, J, Montaigne, M, Parsons, S and Whalan, E 2022, 'Looking under the lamppost or shining a new light: New data for unseen challenges'.
- Hamilton, S, Liu, G and Sainsbury, T 2023, *Early pension withdrawal as stimulus*, Tax and Transfer Policy Institute.
- Harron, K, Dibben, C, Boyd, J, Hjern, A, Azimae, M, Barreto, ML and Goldstein, H 2017, 'Challenges in administrative data linkage for research', *Big Data & Society*, vol. 4, no. 2, pp. 1–17.
- Hashimoto, RE, Brodt, ED, Skelly, AC and Dettori, JR 2014, 'Administrative Database Studies: Goldmine or Goose Chase?', *Evidence-Based Spine-Care Journal*, vol. 5, no. 2, pp. 74–76.
- Hendrickson, L, Bucifal, S, Balaguer, A and Hansell, D 2015, 'The employment dynamics of Australian entrepreneurship'.
- Hindmarsh, G, Laurens, KR, Katz, I, Butler, M, Harris, F, Carr, VJ and Green, MJ 2021, 'Child protection services for children with special healthcare needs: A population record linkage study', *Australian Journal of Social Issues*, vol. 56, no. 2, pp. 223–243.
- Jervelund, SS and De Montgomery, CJ 2020, 'Nordic registry data: value, validity and future', SAGE Publications Ltd STM, *Scandinavian Journal of Public Health*, vol. 48, no. 1, pp. 1–4.
- Korda, RJ, Biddle, N, Lynch, J, Eynstone-Hinkins, J, Soga, K, Banks, E, Priest, N, Moon, L and Blakely, T 2020, 'Education inequalities in adult all-cause mortality: first national data for Australia using linked census and mortality data', *International Journal of Epidemiology*, vol. 49, no. 2, pp. 511–518.
- Laukova, D, Bernard, A and Sigler, T 2022, 'Settlement and Migration Patterns of Immigrants by Visa Class in Australia', in Kawano, M, Kourtit, K, Nijkamp, P and Higano, Y (eds), *Theory and History in Regional Perspective: Essays in Honor of Yasuhiro Sakai*, Springer Nature, Singapore, pp. 165–191.
- Leigh, A 2024, 'Data-driven decisions: the case for randomised policy trials', Address, Treasury Ministers, Oxford University, England.
- Liu, B, Stepien, S, Dobbins, T, Gidding, H, Henry, D, Korda, R, Mills, L, Pearson, S-A, Pratt, N, Vajdic, CM, Welsh, J and Macartney, K 2023, 'Effectiveness of COVID-19 vaccination against COVID-19 specific and all-cause mortality in older Australians: a population based study', *The Lancet Regional Health - Western Pacific*, vol. 40, p.
- McMillan, H and Burns, C 2021, *BLADE for productivity research*, Productivity Commission Staff Working Paper, Canberra.
- Microsoft and Weston, S 2022, *foreach: Provides Foreach Looping Construct*, R Package.
- Mühleisen, H and Raasveldt, M 2024, *duckdb: DBI Package for the DuckDB Database Management System*, R Package.
- NCVER (National Centre for Vocational Education Research) 2023, *From VET to sustainable employment for Aboriginal and Torres Strait Islander peoples*.
- Nicolaou, A, Neumann, Parker, S, Seyani, M, Ryan, N, Gransden, A, Yemm, N and Hope, A 2023, *First Nations People Workforce Analysis*, Jobs and Skills Australia, Canberra.
- OECD (Organisation for Economic Cooperation and Development) 2023, 'Migration and regional productivity: Evidence from individual wages in Australia', *OECD Regional Development Papers*, no. 60.
- Parker, T 2017, 'The DataLab of the Australian Bureau of Statistics', *Australian Economic Review*, vol. 50, no. 4, pp. 478–483.
- PBO (Parliamentary Budget Office) 2021, *Bracket creep and its fiscal impact*, 1.
- PC (Productivity Commission) 1999, *Australia's gambling industries*, Report no. 10, Canberra.
- 2017, *Data Availability and Use*, Report no. 82, Canberra.
- 2018, *Rising inequality? A stocktake of the evidence*, Commission Research Paper, Canberra.
- 2020, *Mental Health*, Report no. 95, Canberra.
- 2024a, *A path to universal early childhood education and care – Vol. 3*, Inquiry report no. 106, Canberra.
- 2024b, *A snapshot of inequality in Australia*, Research paper, Canberra.
- 2024c, *Fairly equal? Economic mobility in Australia*, Research paper.
- Pham, C, Downes, M, Guo, J, Jahan, F, De Silva, S, O'Connor, E, Gray, S, Priest, N and Goldfeld, S 2024, *Measuring vulnerability and disadvantage in early childhood data collections: Phase two*, Centre for Community Child Health, Melbourne.
- R Core Team 2023, *R: A Language and Environment for Statistical Computing*, R Package.
- Rankin, P, Staton, S, Jones, A, Potia, AH, Houen, S, Healey, B and Thorpe, K 2024, *Linking quality and child development early childhood education and care: Technical report*, Australian Education Research Organisation.
- Richardson, N, Cook, I, Crane, N, Dunnington, D, François, R, Keane, J, Moldovan-Grünfeld, D and Ooms, J 2023, *arrow: Integration to 'Apache' 'Arrow'*, R Package.
- Shaw, RJ, Harron, KL, Pescarini, JM, Pinto Junior, EP, Allik, M, Siroky, AN, Campbell, D, Dundas, R, Ichihara, MY, Leyland, AH, Barreto, ML and Katikireddi, SV 2022, 'Biases arising from linked administrative data for epidemiological research: a conceptual framework from registration to analyses', *European Journal of Epidemiology*, vol. 37, no. 12, pp. 1215–1224.
- Smith, G, Noble, M, Anttila, C, Gill, L, Zaidi, A, Wright, G, Dibben, C and Barnes, H 2004, *The value of linked administrative records for longitudinal analysis*, Report to the ESRC National Longitudinal Strategy Committee.
- Suresh, M, Taib, R, Zhao, Y and Jin, W 2019, 'Sharpening the BLADE: Missing Data Imputation Using Supervised Machine Learning', in Liu, J and Bailey, J (eds), *AI 2019: Advances in*

Artificial Intelligence, vol. Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 215–227.

Tang, A, Rankin, P, Staton, S and Thorpe, K 2024, 'Access to high-quality early care and education: Analysis of Australia's national integrated data', *Early Childhood Research Quarterly*, vol. 67, pp. 352–362.

Treasury 2024, *Amending tax cuts to deliver broader cost-of-living relief*.

UNSD (United Nations Statistical Division) 2020, *UNSD - Admin Data Collaborative - Guiding Principles*, <https://unstats.un.org/UNSDWebsite/capacity-development/admin-data/principles> (accessed 6 September 2024).

Varela, P, Husek, N, Williams, T, Maher, R and Kennedy, D 2021, 'The lifetime fiscal impact of the Australian permanent migration program', Treasury Paper.

Welsh, J, Joshy, G, Freeman-Robinson, R, Biddle, N, Banks, E, Schlegel, C, Jordan, K, Gould, P, Kelly, P and Korda, R 2024, 'Understanding the fatal burden of COVID-19 in residential aged care homes in Australia: Using linked data to generate evidence', *International Journal of Population Data Science*, vol. 9, no. 5.

Wickham, H, Girlich, M, Fairbanks, M and Dickerson, R 2023, *dplyr: Data Table Back-End for 'dplyr'*, R Package.

Wright, D 2021, *MADIP – What is it and where is it headed?*, Australian Bureau of Statistics.