Australian Bureau of Statistics

# 2024-25 PLIDA Board Annual Report

# Contents

# Introduction

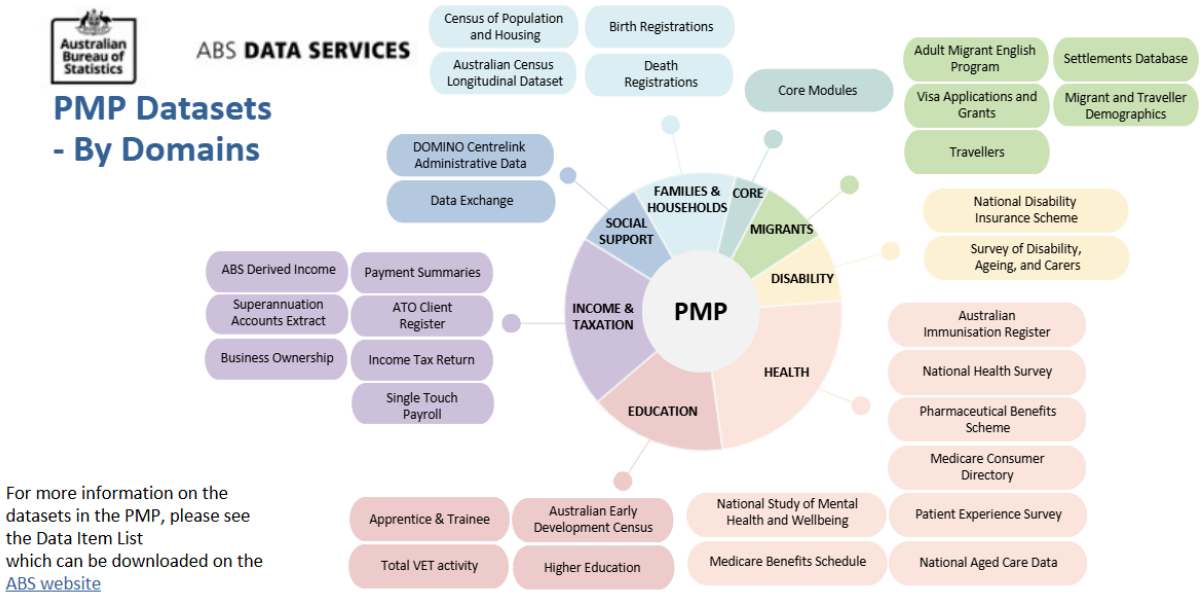
Welcome to the 2024-25 financial year PLIDA Board Annual Report. This Report intends to communicate information on DataLab key metrics, ABS data integration (DI) work program, notable outcomes, client survey satisfaction results and, major investments for product improvement and client services.

## Person Level Integrated Data Asset (PLIDA)

The Person Level Integrated Data Asset (PLIDA) is a secure data asset combining information on health, education, government payments, income and taxation, employment, and population demographics to create a comprehensive picture of Australia over time. There are 37 datasets that are part of the core PLIDA product but the Australian Bureau of Statistics (ABS) has and is linking up to another 84 datasets for specific projects (121 in total). This includes the *General Social Survey for Measuring What Matters* and a lot of State and Territory data and some private sector data. In comparison, there were only 9 datasets in 2017.

PLIDA data is securely held by the ABS with access only made available to approved researchers for approved purposes.

The PLIDA Modular Product (PMP) is a detailed microdata product created from the PLIDA data asset. The PMP contains data from a range of different data custodians and covers several domains.




## Strategic PLIDA updates and PLIDA data expansion

There are several strategic initiatives underway, including:

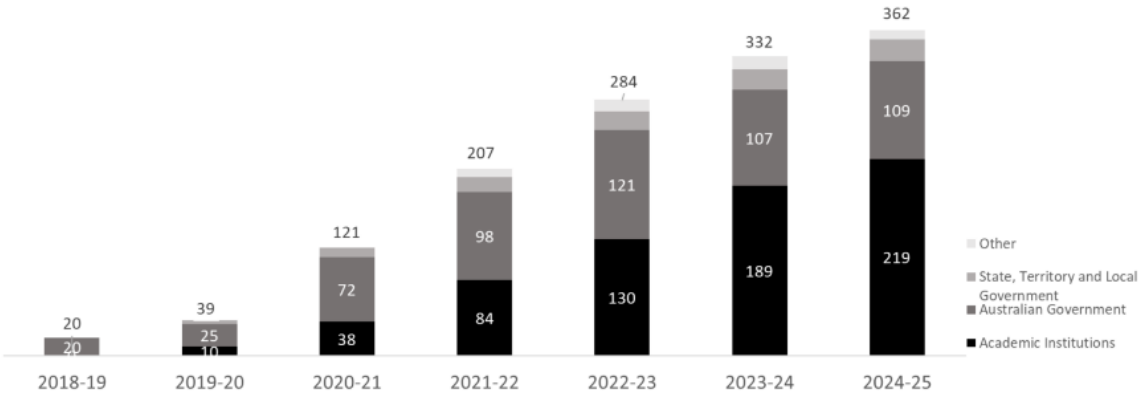
- Discussions with Treasury colleagues on development of housing and wealth data assets.
- Rollout of Single Touch Payroll (STP) Phase Two in collaboration with the Australian Taxation Office (ATO).
- Advancement of the Commonwealth Data Integration Strategy, which emphasizes the value of integrated data for research and policy.
- Department of Health to launch Australia's first health and medical research strategy, looking to leverage integrated data assets like PLIDA.
- Department of Education leading the National Collaborative Research Infrastructure Strategy to improve academic access to integrated data.

- Supporting data access for communities – the National Indigenous Australians Agency (NIAA) and ABS are involved in a proof-of-concept project with Ngurratjuta/Pmara Ntjarra Aboriginal Corporation using PLIDA to derive valuable information about their community.
- Typical integration time ranges from one to three months for person level linkage, depending on the complexity and type of data. Generally, 1 month if identifier is available, around 3 months if there is a need to link names and addresses.
- PLIDA currently has 27 PMP datasets, with up to 62 expected by the end of 2025.

## Historical data based on DataLab

Figure 1 below, shows the count and growth in integrated data projects by organisation type for period 2018-19 to 2024-25.

**Figure 1. Count of integrated data projects by organisation type, 2018-19 to 2024-25**




Notes: Prior to Nov 2021 the old (pre-cloud) DataLab was still in use by clients. Numbers have been sought from the old DataLab, but both start and end dates are unknown. The data should be read as cumulative, for example five projects that are still active have been running since 2016-17 or 2017-18.

## DataLab Key Metrics

Key metrics presented in table 1 show interactions related to the DataLab. The ABS DataLab is a secure, cloud-based research environment provided by the ABS. It allows approved users to access and analyse detailed microdata for statistical and research purposes.

**Table 1: Counts of DataLab Users[1] and DataLab User Organisations[2]**

| Users | 23/24 Counts | 24/25 Counts | Difference (on 2023-24) |
|---|---|---|---|
| Number of unique[3] active users[4] as of 30 June | 220 | 319 | ~45% increase |
| Number of unique active users as of 30 June[5] (cumulative total from 2020) | 2416 (total from 2020 – June 2024) | 2862 (total from 2020 – June 2025) | ~18% increase |
| Number of unique users undertaking DataLab[6] training as of 30 June | 1307 | 1424 | ~9% increase |
| Total Organisations in the DataLab | 98 | 71 | ~28% decrease |

| | | | |
|---|---|---|---|
| Number of users organisations as of 30 June (cumulative total from 2020) | 173 (total from 2020 – June 2024) | 178 (total from 2020 – June 2025) | ~3% increase |

Footnote: [1]An individual who has been approved to access the ABS DataLab for the purpose of conducting statistical or research analysis using detailed microdata. [2] A DataLab user organisation is formally approved by the ABS to allow its approved staff or affiliates to access and use the ABS DataLab. [3] Unique users are only counted once to prevent duplication of the same user across multiple projects within the same organisation. Organisations in the DataLab can have multiple projects often with the same User appearing in more than one project. [4] Users logged into the DataLab and actively running a virtual machine (VM) for their analyses. Includes only 'Analysts' and exclude 'Discussants'. [5] Figures do not reflect users prior to migration from on-premises to cloud and only include 'Analysts'. [6] Count includes 553 users who attended refresher training in FY 24/25 and 432 users who attended refresher training in FY 23/24.

**Table 2: Top ten User Organisations (FY 2024/25):**

Table 2 presents metrics for the top 10 active & inactive, as well as only active DataLab user Organisations, from July 2024 to June 2025. Figures refer to users only counted once to prevent duplication of the same user across multiple projects within the same organisation.

| Rank | User Organisations and User Counts | |
|---|---|---|
| | **Active & Inactive[1] DataLab user organisations with user counts** | **Active DataLab user organisations with user counts** |
| 1 | Australian Bureau of Statistics - 78 | Australian Bureau of Statistics - 57 |
| 2 | Department of Employment and Workplace Relations - 37 | Department of Employment and Workplace Relations - 28 |
| 3 | Department of Education - 24 | Australian National University - 16 |
| 4 | The Treasury - 19 | University of Sydney - 15<br>The Treasury - 15 |
| 5 | Australian Institute of Health and Welfare - 17<br>Australian National University - 17 | Australian Institute of Health and Welfare - 14 |
| 6 | University of Sydney - 16 | Department of Education - 13<br>National Disability Insurance Agency - 13 |
| 7 | National Disability Insurance Agency - 13 | University of Melbourne - 11 |
| 8 | University of Melbourne - 12 | Monash University - 10 |
| 9 | Monash University - 11 | Department of Industry, Science and Resources - 6<br>University of Adelaide - 6 |

| | | |
|---|---|---|
| 10 | Department of Industry, Science and Resources - 10 | Reserve Bank of Australia - 5<br>University of New South Wales - 5<br>University of Chicago - 5 |

Footnote: [1]Inactive counts include researchers who may be trained and approved by the ABS but not actively running a virtual machine for their analyses.

**Table 3: Counts of open and closed DataLab projects**

| Project Counts | 23/24 Count | 24/25 Count | Difference (on 2023-24) |
|---|---|---|---|
| Total projects – (Archived, Closed & Open) | 209 | 264 | ~26% increase |
| Archived & Closed[1] projects | 102 | 150 | ~47% increase |
| Open projects | 107 | 114 | ~7% increase |

Footnote: [1] Please note log/record of closed projects are from March to June 2025, which is 15 closed projects.

**Table 4: DataLab unique Outputs[1]**

| Outputs | 23/24 Count | 24/25 Count | Difference (on 2023-24) |
|---|---|---|---|
| **Total** | 2,555 | 2,185 | ~14% decrease |

Footnote: [1]Counts of unique output requests. Excludes output request reviews, input and transfer requests.

**Table 5: Top 10 projects by output volume[1] (all time-for period Nov 2022-July 2025)**

| Rank | Projects | Number of output requests |
|---|---|---|
| 1 | 2021010-Health Understanding health policy impacts on socio-demographic cohorts | 681 |
| 2 | 2021097-Micro heterogeneity and macroeconomic performance: before and after the virus | 331 |
| 3 | 2018019-Disability and social determinants of health in Australia | 174 |
| 4 | 2021089-Improving Australians health understanding relationships between health and other factors | 157 |
| 5 | 2020098 - Skills Tracker | 149 |
| 6 | 2021056 - Education, Skills and Employment National Data Asset (ESENDA) – Phase 1 | 139 |

| | | |
|---|---|---|
| 7 | 2020z008 - Understanding spare capacity in the labour market | 138 |
| 8 | 2022020 - Census Futures research for the 2026 Census of Population and Housing | 118 |
| 9 | 2023127 - Competition Review | 108 |
| 10 | 2021062 - Enhancing household expenditure data through MADIP | 92 |

Footnote: [1] based on counts of unique requests, which excludes any reviews conducted.

**Table 6: Top 10 projects by output volume[1] (FY 2024/25)**

| Rank | Projects | Number of output requests |
|---|---|---|
| 1 | 2021010-Health Understanding health policy impacts on socio-demographic cohorts | 140 |
| 2 | 2021097-Micro heterogeneity and macroeconomic performance: before and after the virus | 130 |
| 3 | 2021089-Improving Australians health understanding relationships between health and other factors | 109 |
| 4 | 2023127-Competition Review | 91 |
| 5 | 2020098 - Skills Tracker | 47 |
| 6 | 2021056 - Education, Skills and Employment National Data Asset (ESENDA) – Phase 1 | 45 |
| 7 | 2022020 - Census Futures research for the 2026 Census of Population and Housing | 43 |
| 8 | 2019z008 - Ongoing research for annual wage reviews | 37 |
| 9 | 2020z008 - Understanding spare capacity in the labour market | 37 |
| 10 | 2020071 - Treasury Real-Time Labour Market Tracker | 35 |

Footnote: [1] based on counts of unique requests, which excludes any reviews conducted.

**Table 7: Data Products[1]**

| File Directories/Data Products | Count | Accessed/Available |
|---|---|---|
| Accessed | 2452 | Number of product file directories[2] (not specific products) accessed by Users for FY 2024/25 |

| | | |
|---|---|---|
| Available | 1988 | Total number of data products in the DataLab (data does not contain dates for FY) |

Footnote: [1] Data products are the specific microdata files or data sets used for analysis. [2] Product file directories are the folder directories that contain the Data Products and the number of unique accesses.

## Data Integration Budget 2024-25

For 2024-25 the ABS data integration program cost $17.2m, of which the majority:

- $7.4m is funded by the Managing Australia's Essential Statistics base investment by the Government
- $5.9m is funded by other government investments, New Policy Proposals (NPPs) and ABS appropriation, and
- $3.9m is funded by client revenue based on the ABS pricing. Of this $3.9m,
    - $1.6m is for data linkage work. Data linkage is fully cost-recovered, and the underlying cost model is reviewed biannually. With ABS funding at full capacity, updates to new datasets, including the ones listed above, need an ongoing funding source.
    - $2.3m is for DataLab access. DataLab access is currently heavily subsidised by ABS, and the underlying cost model is reviewed annually.

Of the 420 DataLab projects, around 63% are university led (266). Of the $17.2m cost of the ABS DI Program, the university sector currently contributes $0.55m or ~3%, almost entirely via the Universities Australia (UA) Agreement. There remain more projects operating under the UA agreement than the funding from it supports. In the 2025 calendar year, we expect funding gap for academic projects under the UA Agreement to be $700k. We continue to operate 2 projects out, 1 project in model with a goal of reducing to 180 projects in 2026 and are negotiating an increase to the UA cost. This will result in the UA agreement almost funding all academic research (with ~$50k deficit). We will break even in 2027 and revert to a 1 out, 1 in model. The growth and maintenance of academic project involvement in PLIDA requires a substantial investment from the sector.

## Notable Outcomes

Notable outcomes highlight the broader implications of researchers' access to the data set.

a. Supporting the collection and production of official statistics using PLIDA

   The ABS is continuing to expand its utilisation of administrative data sources from PLIDA for official statistics, particularly in household survey collections. Over the past year, data in PLIDA has been used to enable more efficient sampling design, supplement survey responses, support output weighting and inform quality assurance activities. Increasingly, the final survey data is being incorporated into PMPs where it can be utilised for a wide range of approved research projects. This work is expected to reduce collection costs, improve representation in survey populations, deliver more accurate outputs and enrich the data available for decision making.

b. *StatsLab* to support ABS production of official statistics from integrated data products.

   The PLIDA Board endorsed the concept of *StatsLab*, intended for internal ABS use to make more effective use of PLIDA data in the production of official statistics. *StatsLab* use similar technologies to DataLab and is part of the ABS data modernisation work program. *StatsLab* is accessible only to ABS staff, who sign lifelong undertakings under the *Census and Statistics*

*Act 1905*. It is currently in an Alpha testing phase, with Census Operational readiness Exercise (ORE) analysis currently underway.

c. DataLab annual fees increased by 10% from July 2025 because of underlying and cumulative cybersecurity, cloud and labour cost pressures. Data Integration prices have not changed due to efficiency improvements in the linkage process. The ABS supports around 2,000 researchers across ~420 projects to access integrated and microdata products via the DataLab, the key platform for safely accessing integrated data for research, program evaluation and policy development. DataLab projects involve governance and data operations costs that have previously been heavily subsidised by the ABS, to support the sustainability of the Data Integration program, these must be cost-recovered.

d. Access outside of Australia has been approved by the ABS Executive Board, along with a pricing structure for the service. As of the start of October, we have 17 researchers across 4 different countries, representing 6 organisations on 13 projects who are accessing DataLab overseas.

e. Approval was obtained and a technical solution for virtual collaboration (screen sharing) in the DataLab has been implemented. LabLink is a new video conferencing tool developed by the ABS for SEAD, now available in its trial (BETA) phase to all users. It is designed to help researchers collaborate securely within their Virtual Machines (VMs), even when working from different physical locations. There are no additional costs involved for projects using LabLink.

f. The *ABS Data for Evaluation Projects* policy was endorsed, formalising operational arrangements and legislative mechanisms for releasing data to support program and service evaluation activities. Three projects have been approved under these arrangements: ACT justice data, Mission Australia housing and homelessness services data, and the Home Interaction Program for Parents and Youngsters (HIPPY) program data.

g. The ABS engaged independent privacy consultant Maddocks, on behalf of the PLIDA Board, to conduct the 2024-25 PLIDA Privacy Impact Assessments (PIA) Update. The PLIDA Board has agreed with all the recommendations from the 2024-25 PLIDA PIA Update and their response provides a summary of recommendations together with how the recommendations will be implemented. The recommendations focus on actions to advance privacy best practice and enhance existing governance processes. The ABS is working with the PLIDA Board to implement these recommendations by December 2025.

h. The Board supported 8 projects that were selected for the pilot trial for managing *'Enduring Multipurpose Programs (EMPs)'* for research and analysis.

i. A new webpage on the ABS website dedicated to the PLIDA Board was published, that houses meeting notes available for reference. This contributes to the transparency of PLIDA operations, including regular updates to the ABS website with reference to PLIDA.

j. The Board endorsed the updated *Terms of Reference* (TOR). Update includes a refresh of member details and references the responsibility for members to 'apply a whole-of-Commonwealth perspective to maximise data utilisation in the interests of the Australian community'.

## 2025 DataLab and Data Integration User Experience Survey Results

This is an annual survey and was run for the third time between 19 May to 6 June 2025. Participants in this survey are DataLab project leads and active DataLab users. 217 (13.2%) responses were received, out of which 75% were PLIDA users. This was **up 1.5% from 2024** where 182 (11.7%) responses were received.

Results are in the process of being finalised and further details will be presented at the October 2025 PLIDA Board meeting. Below are **high level outcomes** from the Survey:

- A mix of new and experienced users, out of which 24% of participants have used DataLab for less than 1 year
- Users of all products but 82% of respondents primarily used PLIDA modular products
- Sector composition of respondents included Commonwealth 58%, University/academic 35%, State 6%, Public and Policy Research Institute (PPRI) 2%
- Around 75% of the survey respondents had used PLIDA for a research project. The majority of these were satisfied with PLIDA.
- 79% indicated that they were satisfied or very satisfied with their overall experience using PLIDA, which went down by **1% from last year.**
- Most reported that PLIDA met their research needs, and there were no standout barriers to their use.
- Users were also really satisfied with the core modules.
- 69% indicated they were satisfied or very satisfied with their overall experience using the DataLab, which went **up 3% from last year.**
- 77% indicated that they were satisfied or very satisfied with their overall experience of the custom DI service – **up 11% from last year.**

Results are being finalised, but **Initial takeaways** highlight that:

- The majority are satisfied with DataLab and Data Integration (DI) services. Overall satisfaction ratings across key areas including DataLab system, DataLab Onboarding and Training, PLIDA and Business Longitudinal BLADE appear to be consistent with previous results.
- The level of satisfaction with myDATA was lower than others, free-text comments indicate a high level of frustration and some discontent.
- Results and free-text comments regarding Output vetting and DataLab charging are in line with expectations and highlight areas for improvement.

## Major Investments for product improvement and client services

There are several investments that the ABS is making to reduce costs, improve the product offering, and deliver more timely and transparent services to its clients:

a. myDATA client relationship management tool commenced being used in June 2024. Data Integration Services, Digital Services and Technology services branches have jointly initiated a short-term project to scope improvements for the myDATA User experience.

b. The Economic Data Integration Team (EDIT) has commenced the migration of BLADE processes to Australian National Data Integration Infrastructure (ANDII), focusing on core processes that are common across multiple stages of the BLADE integration process. It is expected that the migration of PLIDA processes to ANDII will be complete by the end of 2025 with support from Statistical Infrastructure Division.

c. The streamlining of data integration and program management in June 2024 saw improvement in the Data Integration Pipeline processes, that has created visibility, scheduling and capacity planning.

d. New Core modules were delivered in 2024-25 financial year, with plans to update and further deliver additional new Core modules in the 2025-26 financial year.

e. Data linkage operations have been migrated to the cloud.

f. The Board approved the development of PLIDA *TableBuilder* products. The products will enable access to integrated data for those without the capability or funding to access microdata in the DataLab. *TableBuilder* products have been requested by clients and Board members for some years. These products allow for instant (seconds) production of tables of counts from PLIDA assets.

g. Please advise if you would like any further updates on major investments for product improvement and client services.as it will help inform future development of this part of the report.


www.abs.gov.au