



Maddocks

Australian Bureau of Statistics

**Personal Level Integrated Data Asset Update
2024-25**

Consultation Report

4 April 2025

Contents

- Part A Overview 3**
- 1. Introduction 3
- 2. Approach to stakeholder consultations..... 3

- Part B Summary of stakeholder feedback 5**
- 3. Introduction 5
- 4. Inclusion of new and expanded data types in PLIDA 5
- 5. Creation of expanded PLIDA outputs 9
- 6. New or expanded data handling practices for PLIDA..... 10

- Attachment 1 Stakeholders consulted for the 2024/25 PLIDA Update PIA..... 12**

- Attachment 2 Copy of material provided to stakeholders – Description of PIA issues..... 13**
- 1. Overview of the Person Level Integrated Data Asset (PLIDA)..... 13
- 2. Overview of the 2024-25 PLIDA Update 15
- 3. Including new or expanded data in PLIDA 16

- Attachment 3 Copy of material provided to stakeholders – Consultation slides 24**

Part A Overview

1. Introduction

- 1.1 The Person Level Integrated Data Asset (**PLIDA**), formerly known as the Multi-Agency Data Integration Project (**MADIP**), is a secure, enduring, person-based research data asset that combines broad sets of information about Australian populations. PLIDA creates a comprehensive picture of Australian residents over time, and facilitates the use and re-use of public data for statistical analysis and research purposes.
- 1.2 The Australian Bureau of Statistics (**ABS**) is the accredited integrating authority for PLIDA and is responsible for combining the datasets in PLIDA, for providing access to PLIDA data to authorised users for approved research projects, for ensuring the security of data contained within PLIDA, and ensuring that all research results and other outputs from the use of PLIDA data are done in a manner that is unlikely to enable the identification of a particular person.
- 1.3 The ABS has undertaken three major privacy impact assessment (**PIA**) processes in relation to PLIDA to date (in 2018, 2019 and 2022), with each subsequent PIA process considering planned changes or updates to PLIDA since the previous process. Various project-specific PIAs involving PLIDA data have also been conducted in between the major PIA processes during this time.
- 1.4 The ABS has engaged Maddocks to conduct a further PIA update process (**the 2024-25 PIA Update**) to:
 - 1.4.1 consolidate consideration of the privacy impacts of some changes or updates to PLIDA since 2022 (taking into account the previous separate PIA and other privacy consideration that has already been given to address those changes or updates);
 - 1.4.2 consider some future planned updates or changes to PLIDA to ensure continued compliance with the Privacy Act, the Australian Privacy Principles (**APPs**), and privacy best practice; and
 - 1.4.3 facilitate the ABS continuing to take a 'privacy by design' approach to PLIDA.
- 1.5 Throughout October 2024, the ABS conducted an extensive consultation process with identified stakeholders.
- 1.6 This Consultation Report provides a summary of the feedback and concerns raised in the consultation sessions held with stakeholders. The consultations will inform the privacy risk analysis and recommendations for the 2024-25 PIA Update being conducted by Maddocks.

2. Approach to stakeholder consultations

- 2.1 The objectives of the consultation processes were to:
 - 2.1.1 invite stakeholder feedback and insights in relation to each of the issues within the scope of the 2024-25 PIA Update (**PIA Topics**);
 - 2.1.2 inform stakeholders about PLIDA, including its current privacy practices and protections; and
 - 2.1.3 listen to issues and concerns stakeholders may have about the PIA Topics.

- 2.2 The stakeholders consulted included representatives of Commonwealth agencies on the PLIDA Board (who are data custodians), university academics, government officials, the Office of the Australian Information Commissioner (**OAIC**) and its State and Territory regulator counterparts, data custodians from States and Territories, charities, peak bodies and representatives of Aboriginal and/or Torres Strait organisations. **Attachment 1** at the end of this report sets out the full list of stakeholder organisations that participated in the consultations.
- 2.3 Stakeholders were invited to participate in consultation sessions based on their:
- 2.3.1 involvement in current, or prior Australian Government data integration activities and projects;
 - 2.3.2 special interest or expertise, such as advocating about privacy and data; and
 - 2.3.3 representation of a key sector of Australian society that is likely to be impacted by PLIDA and its outputs, such as Aboriginal and/or Torres Strait Islander Australians.
- 2.4 Whilst stakeholders had the opportunity during the consultation sessions to raise any concerns that they had about the PIA Topics, they were also invited to provide subsequent written submissions to the ABS. A number of stakeholders took up this opportunity, and all such feedback received was provided by the ABS to Maddocks.
- 2.5 The consultation sessions were conducted virtually in five sessions of up to two hours each with different stakeholder groups. Participants were provided with a description of the PIA Topics, at least one week in advance of each session. A copy of the material provided to participants is in **Attachment 2**.
- 2.6 The sessions were run by the ABS and supported by Maddocks. A copy of the slide pack presented at the sessions is in **Attachment 3**.
- 2.7 Consultation sessions covered an overview of PLIDA and the underlying infrastructure, how data is kept safe, and how ethics and specific privacy considerations are managed for PLIDA. Information about each of the PIA Topics was then provided, and stakeholders were invited to make comments, ask questions or provide feedback on what had been discussed. Questions were also encouraged throughout each session.

Part B Summary of stakeholder feedback

3. Introduction

- 3.1 This Part sets out a thematic summary of the views, opinions and key discussion points expressed by stakeholders throughout the consultation process.
- 3.2 Maddocks has prepared this summary without attributing any particular opinion to any individual or stakeholder. However, some comments have been attributed to the type of stakeholder who provided the view (e.g. a Government entity), to provide context.
- 3.3 The aim of the consultations was to understand stakeholder views on the purpose and potential benefits associated with the PIA Topics, as well as to highlight any areas of concern amongst the stakeholders. These concerns could then be considered and addressed in the 2024-25 PIA Update report.
- 3.4 Accordingly, this Consultation Report does not set out all of the existing mitigation strategies that are currently in place, or which the ABS intends will be implemented, which might be relevant to the PIA Topics. Rather, this Consultation Report is simply intended to serve as a record of the issues that were discussed during the consultation sessions, including how the issues were understood by the participants. This Consultation Report should therefore be read in conjunction with the 2024-25 PIA Update report.
- 3.5 The summary of stakeholder feedback below details key issues that were raised during the consultation sessions, for each PIA Topic.

4. Inclusion of new and expanded data types in PLIDA

- 4.1 During the consultation sessions stakeholders were told about the changes in the types of data that are to be considered as part of the PIA

PIA Topic 1: Linking high level administrative crime and justice data to PLIDA

- 4.2 Generally, stakeholders, particularly those who were data custodians, were comfortable with the application of the standard PLIDA practices to govern inclusion of high level administrative crime and justice data in PLIDA, and its subsequent use for research projects - particularly taking into account the current role of the relevant data custodian(s) in approving that inclusion and use.
- 4.3 Stakeholders also generally considered that it was appropriate that the proposed high level data to be linked did not include more detailed crime and justice data about the relevant data subjects, and for specific privacy consideration to continue to be given to PLIDA projects that would involve the inclusion of any more detailed crime and justice data. Although one stakeholder noted that while having data at such a high level may reduce its overall sensitivity and privacy risk, it may also impact the data's usefulness for research projects, particularly where detailed data would be useful for longitudinal studies of people who had been affected by the justice system.

- 4.4 Most stakeholders, particularly researchers, spoke positively on the public benefit accrued by researchers having access to high-quality data of this nature for policy development, noting that the relevant cohort of data subjects is heavily impacted by government policy, especially during their resettlement into society. Despite being high level, it was felt that the administrative crime and justice data would assist in providing visibility for a population group that might not otherwise be easily ascertained from other types of data (for example, individuals in prison would not be accessing healthcare services via Medicare, so they would not be represented in MBS/PBS data).
- 4.5 Some stakeholders noted the interaction between data ethics and privacy. One stakeholder commented on the emerging public expectations of data autonomy (i.e. people having control of their data), particularly if it is being used for a purpose for which it was not initially disclosed to a government agency. While the administrative crime and justice data will be high level, stakeholders noted that the sensitive nature of the data meant that the individuals to whom it belongs may expect to have a level of control over it, particularly where consent is not usually the authority relied on by the data custodian for including that data in PLIDA. There was some general discussion about the need for data custodians to communicate to individuals at the point of collection about how that their personal information may be used for research purposes (via PLIDA).
- 4.6 Some stakeholders also raised the possibility of an enhanced risk of re-identification when high level crime and justice data is linked to other data items for specific research projects, particularly given the vulnerability of the relevant cohort and the potential harms that could be caused if reidentification was to occur. Even if reidentification is unlikely, stakeholders again emphasised the importance of the public being made aware of how their data might be used, particularly if that use is not consistent with the purpose for which it was originally provided.
- 4.7 Stakeholders also raised the need to be clear about the quality of this type of data, if inferences about a data subject's relationship and interactions with the justice system, which is an inherently sensitive topic, were going to be made using the high level administrative crime and justice data. Stakeholders emphasised the need to ensure that people applying to access the administrative crime and justice data should continue to be required to consult with the relevant data custodian(s), to ensure that it is being used and interpreted correctly.
- 4.8 Some stakeholders noted the importance of conducting additional privacy impact assessments before including any state-based crime and justice data in PLIDA. However, they seemed generally comfortable with the ABS' approach of leveraging any previously performed specific privacy work on similar data (e.g. NSW Their Futures Matter) to ensure that conceptually all domains are covered.
- 4.9 Some stakeholders noted the particular sensitivity of crime and justice issues within the Aboriginal and/or Torres Strait Islander community, noting concerns around the high level administrative crime and justice data being used to further deficit narratives, especially if interpreted without an Aboriginal and/or Torres Strait Islander perspective.

PIA Topic 2: Linking 'not for profit' (NFP) organisational data to PLIDA

- 4.10 Stakeholders appeared split on the benefits of including NFP data from NFP organisations in PLIDA.
- 4.11 Generally, stakeholders who were existing data custodians were supportive of NFP organisations providing data directly to the ABS. This was on the basis that NFP data has the potential to deliver the same public benefit as non-NFP data, should not be considered differently from non-NFP data, and that the expanded data provision would be captured in a contract or data sharing arrangement with the NFP organisation. One group of stakeholders noted that NFP data would be beneficial for research projects to understanding wellbeing (and therefore entrenched disadvantage) from a holistic perspective, given that NFP organisations tended to use different metrics and tools to measure the success of the

services they deliver. Other stakeholders noted the importance of NFP data to inform social impact investing.

- 4.12 Other groups of stakeholders were less convinced about the additional benefits of receiving and linking NFP data from NFP organisations, because there are already avenues in place for such data to be received as administrative data received from government data custodians, and that typically the data obtained via these avenues was sufficiently detailed. Such stakeholders sought clarification about the additional benefits that would arise from obtaining NFP data straight from the NFP organisation rather than the government agency.
- 4.13 Another stakeholder noted that, on an ongoing basis, the ABS should consider whether increasing the type and volume of NFP data is proportional to benefits that arise from including such datasets in PLIDA.
- 4.14 Stakeholders commented broadly on some issues extending beyond privacy risks, such as the potential for this type of data to have a commercial-in-confidence quality, and for some NFP organisations to still have commercial purposes or interests in connection with the NFP data (for example, a drive to secure further funding or clients, and/or to prevent competitors from securing them).
- 4.15 Stakeholders observed that the governance arrangements around PLIDA would need to be carefully considered if large NFP organisations were to become data custodians, with rights under the current PLIDA governance framework to restrict access to the data they provide (which they might do to prevent research findings that could be contrary to the NFP organisation's commercial interests). One stakeholder noted that it would not be appropriate for NFP organisations' commercial interests to undermine the ability for PLIDA to be used to derive research insights of public benefit, or for an NFP organisation to access PLIDA data for commercial purposes. Another stakeholder considered that operationally allowing NFP data into PLIDA would make it more difficult to refuse access to PLIDA data by NFP organisations.
- 4.16 Some stakeholders suggested that some NFP organisations might struggle to be effective data custodians – for example, because their data governance, management and vendor arrangements may not be as sophisticated as government data custodians, and they would need to consider PLIDA research proposals resulting in higher administrative burdens and costs for NFP organisations. It was suggested that in some cases, dealing with the NFP organisation as data custodian may be less efficient than if the data was received from government data custodians.
- 4.17 Stakeholders also raised other potential risks associated with a reduced privacy maturity. For example, risks of NFP data being provided to ABS from cloud services or systems that are not certified as secure, or the NFP data being structured in a way that involved an enhanced reidentification risk. It was noted that receiving NFP data from NFP organisations would be considered on a case-by-case basis in accordance with the usual PLIDA processes, and that NFP organisations would not be compelled to provide NFP data to ABS.¹
- 4.18 Stakeholders also queried whether the quality of data provided by NFP organisations may affect the ability for linkage with other PLIDA data, or appropriate use in research projects. It was felt that some NFP data may be of lesser quality or reliability than data provided by other data custodians, especially noting that it may not be subject to the same record-keeping and accuracy standards as government organisations. This may create additional work for the ABS if it needed sanitising or structuring. Other stakeholders noted that some NFP organisations may have less or compromised technical expertise or data literacy (which would affect the quality of NFP data). However, it was noted that these issues should be resolved by the ABS during the PLIDA data cleansing processes.

¹ This reflects the usual position for data custodians under the PLIDA governance framework, noting that data might still be *required* to be produced under the *Census and Statistics Act 1905* (Cth).

PIA Topic 3: 'Sensitive' and 'inherently sensitive' survey data

- 4.19 All stakeholders were generally supportive of the standard PLIDA approach of conducting a PIA before including any survey data containing 'sensitive information' or 'inherently sensitive' data items into PLIDA, particularly if survey respondents who identified as part of marginalised population groups (about whom 'sensitive information' is more likely to be collected) would be able to provide their views as part of the PIA process.
- 4.20 Stakeholders generally agreed on the importance of capturing 'sensitive' and 'inherently sensitive' survey data to create a holistically accurate view of the population. One stakeholder noted that survey data containing ethnicity and sexual orientation data alluded to specific populations being correctly 'counted', which would support the data custodians' abilities to plan, resource and use evidence-based advocacy for appropriate service delivery. However, there was discussion about how to maximise the data's value for research, with the desire to give people autonomy over their 'sensitive' or 'inherently sensitive' information. Implementing consent mechanisms (thereby allowing people to opt-out of the surveys of a sensitive nature) might skew the data, making the data less valuable for research and statistical purposes.
- 4.21 There was also discussion about the difficulties of defining 'inherently sensitive information' and how the definition might vary over time, or when combined with other variables. One stakeholder suggested that, ideally, a process should be built into survey responses, potentially as part of the survey design or pilot testing process for survey respondents to define their 'inherently sensitive' information.
- 4.22 Some stakeholders expressed support for data custodians continuing to be consulted about whether it was appropriate to link the survey data at the enduring asset level, and for that survey data to be used for specific research projects (including whether it was necessary to insist on additional ethics approvals to be obtained for particular projects).
- 4.23 Other stakeholders discussed the potential sensitivity of survey responses, and the appropriateness of using that data for the secondary purpose of linking it to PLIDA, noting that this may be a secondary purpose beyond the reasonable contemplation of the survey respondents. However, it was noted that the use of properly deidentified data for PLIDA research projects posed less of a risk to survey respondents.²
- 4.24 One stakeholder noted difficulties in providing a view on the topic because it was not clear whether the inclusion of survey data would require reidentification before it could be useful, which would violate the separation principle underlying PLIDA's operation.
- 4.25 Generally, stakeholders did not think that there were any types of survey data for which a 'blanket ban' would always be appropriate, and that the use of any data for research projects should be assessed on a case-by-case basis.

² During the consultation sessions, some stakeholders noted terminology issues, suggesting that 'deidentified' or 'anonymised' (in the context of describing the treatment of data in PLIDA) should not be used, but rather the term 'pseudonymised' would be more appropriate.

5. Creation of expanded PLIDA outputs

- 5.1 During the consultation sessions stakeholders were told about the creation of some new or expanded outputs using the data in PLIDA, which would be considered as part of the 2024-25 PIA Update.

PIA Topic 4: Creation of new PLIDA modules, including to support the Life Course Data Initiative (LCDI)

- 5.2 Stakeholders were generally supportive about the creation of the new standard enduring datasets (known as **Core Modules**).
- 5.3 In relation to the Core Module – Indigenous Status, stakeholders discussed concerns surrounding development of the module itself, noting that the new module provides alternative approaches that can be used instead of the “ever identified” approach of the old module. Stakeholders discussed the risk of creating inconsistent data, if users would be deriving their own Indigenous status from across several datasets rather than relying on an ABS-provided single derived status. However, stakeholders also discussed the benefits of having the option to use a pseudo-code to reconstruct the “ever identified” status from the old module, or sharing new methods of derivation with the research community.
- 5.4 Stakeholders also discussed the benefits of having cultural governance and testing (in the form of a Cultural Review Panel) as a component of approving access to Aboriginal and/or Torres Strait Islander data for research projects, in place of formal ethics approval.
- 5.5 Stakeholders were also supportive of the development of new family and household structures as part of the LCDI³, noting that building a comprehensive household picture could be of great public benefit. Stakeholders, in particular those who were researchers, noted the operational benefit of having child-centric data available for modelling, given the dependency on this variable in research about households.
- 5.6 Stakeholders (who were data custodians) highlighted that it would be important to clearly define ‘households’ and ‘families’, particularly if those definitions are to change over time, if where certain data about members of families and households may not be included in the relevant PLIDA data, or where different government programs may define the concepts differently. For data subjects from an Aboriginal and/or Torres Strait Islander background, the concept of ‘family’ may not be defined in biological terms, which may introduce complexities in accurately developing the family structures for the LCDI. Some stakeholders voiced concerns about whether the Life Course Dataset (and PLIDA) should be restructured to accommodate extended relationships of data subjects, and noted that it was difficult to answer questions about the value of the LCDI household and family structures, when the answer was centrally dependent on the specific research questions being asked.
- 5.7 Some stakeholders noted a heightened risk of reidentification amongst smaller geographies, including amongst Aboriginal and Torres Strait Islander communities.
- 5.8 Several groups of stakeholders also raised the risk of inaccurate data being provided by particular communities, which will then be captured in the Core Modules and the Life Course Dataset. For example, underreporting of household inhabitants, a prevalent issue in Aboriginal and/or Torres Strait Islander communities due to fear of consequences from housing authorities, may affect the data’s accuracy.

³ Although not a privacy issue, it was suggested that the ABS may wish to consider changing the name of the LCDI, given that the Murdoch Children’s Research Institute had developed a ‘LifeCourse Initiative’ in 2013, with more than 25 longitudinal cohorts to date (see <https://www.mcric.edu.au/research/research-areas/population-health/lifecourse>).

6. New or expanded data handling practices for PLIDA

- 6.1 During the consultation sessions, stakeholders were told about the potential implementation of new or changed data handling practices for PLIDA.

PIA Topic 5: Expanding the use of PLIDA to support statistical production (including the Census and ABS Household Surveys)

- 6.2 Stakeholders that expressed a view on this PIA Topic were very supportive of the proposed use, indicating that it was a useful initiative that was already being implemented overseas.

PIA Topic 6: Combining the 'librarian' and 'linker' roles

- 6.3 Stakeholders were generally supportive of this functional change, with some emphasising that, as long as the data separation principle was maintained, the proposed change did not cause any concern.

PIA Topic 7: Move from Secure Data Integration Environment (SDIE) to the Australian National Data Integration Infrastructure (ANDII ICT system) for PLIDA data integration

- 6.4 Stakeholders, particularly those with practical experience using the ANDII ICT system (in the context of the NDDA), were supportive of the efficiencies gained by transitioning from the SDIE into the ANDII ICT system.
- 6.5 Some stakeholders seemed to incorrectly conflate the ANDII ICT system with its use for the NDDA, which caused confusion around whether the legislative basis for collection of PLIDA data was changing. (For clarity, the ABS will continue to collect and use PLIDA data under the existing PLIDA governance arrangements and the *Census and Statistics Act 1905* (Cth), and it is not currently proposed to use other laws such as the *Data Availability and Transparency Act 2022* (Cth).
- 6.6 The ABS explained that there would only be a change to the ABS controlled ICT environment that will be used to process PLIDA data (i.e., this change would not affect how data is used for PLIDA). If this change is implemented, data for use in PLIDA will be stored separately to data for use in the NDDA.

PIA Topic 8: Re-use of PLIDA data for the National Disability Data Asset (NDDA)

- 6.7 As mentioned in PIA Topic 7, there was some confusion about NDDA operating separately from PLIDA, which caused some concern for stakeholders who were data custodians about potentially having reduced control over the data shared with the ABS. The ABS explained that the processes for permitting data custodians to agree to the further use of PLIDA data for NDDA purposes would not affect the existing PLIDA governance arrangements.
- 6.8 It was explained that the change would only affect data custodians that had previously shared data with the ABS for PLIDA, and had then separately agreed to permit use that data for the NDDA (under the NDDA governance framework). It was noted that the reuse of PLIDA data for the NDDA would be covered by a data sharing agreement with the relevant data custodian under the NDDA governance arrangements.

PIA Topic 9: Access to DataLab by international researchers

- 6.9 Some stakeholders discussed the various measures which may be implemented before providing DataLab access to internationally-based researchers. Stakeholders agreed that the measures should mitigate security risks associated with disclosing data overseas, while minimising any impacts to the data's utility for research projects, particularly for international researchers collaborating on Australian projects with local researchers.
- 6.10 Some stakeholders noted the jurisdictional challenges around enforcing Australian privacy laws for improper use of data. This was on the basis of legal accountability, whereby investigation or enforcement of Australian privacy laws on foreign individuals who are not bound by them would be difficult, and regulatory compliance, if Australian and foreign privacy laws were in conflict with each other.
- 6.11 To mitigate such risks, one stakeholder suggested implementing a residency condition for sensitive PLIDA data (i.e. making it available only to Australian residents), to ensure that data users could be held to Australian privacy law and regulation. Other stakeholders opposed the provision of data access to international researchers, and suggested it would only be appropriate for international researchers to partake in Australian projects in the role of a discussant only, with no actual access to PLIDA data.
- 6.12 Stakeholders also discussed the potential to provide international researchers with synthetic data, but thought that the data's utility might be diminished if treated with this or data suppression techniques.
- 6.13 Stakeholders also noted the risk of foreign researchers misinterpreting or misrepresenting Australian data, in the absence of contextual understanding of Australian policies, demographics or regions. There was some more specific discussion in relation to Aboriginal and/or Torres Strait Islander data, and how granting access to international researchers would be iniquitous in circumstances where local communities struggle with gaining and maintaining data sovereignty – that is, the Aboriginal and/or Torres Strait Islander community would have no input to the governance of data once it was disclosed overseas. Some stakeholders suggested the imposition of stronger access conditions, such as requiring projects to engage or collaborate with Indigenous organisations before they could gain access to Aboriginal and/or Torres Strait Islander data.
- 6.14 Some stakeholders also noted an ongoing concern about the lack of access to PLIDA for Aboriginal and/or Torres Strait Islander community organisations, and raised specific concerns about a model that allowed researchers overseas to access to Aboriginal and/or Torres Strait Islander data, when relevant communities in Australia could not.

Attachment 1 Stakeholders consulted for the 2024/25 PLIDA Update PIA

The list below sets out the entities that attended the consultation sessions for the PIA⁴:

1. Australian Bureau of Statistics (ABS)
2. Australian Taxation Office (ATO)
3. Barang Regional Alliance
4. Department of Education (Education)
5. Department of Employment & Workplace Relations (DEWR)
6. Department of Health and Aged Care (DHAC)
7. Department of Home Affairs (DoHA)
8. Department of Social Services (DSS)
9. Department of Treasury (Treasury)
10. Services Australia
11. South Australian Department of Treasury and Finance (DTF)
12. ACT Government
13. Office of Information Commissioner Queensland
14. Indigenous Data Network
15. Australian Child Rights Taskforce
16. Australian Institute of Health and Welfare (AIHW)
17. Queensland Government Statisticians Office
18. University of Queensland (UQ)
19. Cancer Institute New South Wales (NSW)
20. Queensland Treasury
21. Canteen Australia
22. Office of the Australian Information Commissioner (OAIC)
23. Department of Communities Western Australia (WA)
24. Monash University
25. Australian Climate Service
26. University of Adelaide
27. University of Technology Sydney (UTS)
28. Glen Group
29. ACT Health
30. Murdoch Children's Research Institute
31. Healing Foundation
32. Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS)
33. Kimberley Aboriginal Medical Services (KAMS)

Several other entities were invited to participate in the consultation sessions but were not able to attend.

⁴ A range of individuals attended from the entities listed. The views expressed however by those individuals were not necessarily provided on behalf of their entity.

Attachment 2

Copy of material provided to stakeholders – Description of PIA issues

[The following sets out the contents of the information provided to stakeholders before the consultation sessions.]

1. Overview of the Person Level Integrated Data Asset (PLIDA)

Context

- 1.1 The Person Level Integrated Data Asset (**PLIDA**), formerly known as the Multi-Agency Data Integration Project (**MADIP**), is a secure, enduring, person-based research data asset that combines broad sets of information about Australian citizens, to create a comprehensive picture of Australians over time, and facilitate the use and re-use of public data for statistical analysis and research purposes.
- 1.2 PLIDA allows better use to be made of information that has already been collected, in a data asset that researchers from government, universities, and public policy institutes can use for approved research and statistical analysis projects, including those that inform government decision-making.
- 1.3 The Australian Bureau of Statistics (**ABS**) is the accredited integrating authority for PLIDA and is responsible for:
 - 1.3.1 receiving data from data providers (**Data Custodians**);
 - 1.3.2 combining those datasets in PLIDA and ensuring the security of data contained within PLIDA;
 - 1.3.3 for providing access to PLIDA data to only those who have been authorised (**Approved Researchers**) for approved projects in a secure, virtual access ABS environment (the **Datalab**); and
 - 1.3.4 ensuring that all research results and other outputs from the use of PLIDA data are done in a manner that is unlikely to enable the identification of a particular person.
- 1.4 PLIDA is bound by the constraints of:
 - 1.4.1 the legislation of the Data Custodians that applies to data that they provide to PLIDA and any conditions imposed in data sharing agreements;
 - 1.4.2 the *Census and Statistics Act 1905 (Cth)* (**Census and Statistics Act**);
 - 1.4.3 the *Privacy Act 1988 (Cth)* (**Privacy Act**);
 - 1.4.4 the PLIDA governance framework; and
 - 1.4.5 social licence from, and the community expectations of, Australians.

PLIDA data

- 1.5 PLIDA contains regularly updated datasets that aim to comprehensively cover the Australian population. The data contained within PLIDA is primarily made up of data that is provided to the ABS by Commonwealth Data Custodians (or other entities authorised by Data Custodians to provide that data). If the relevant individuals in those datasets are able to be identified, the datasets will include personal information and sensitive information (as defined in the Privacy Act).

- 1.6 The ABS has built a central linkage infrastructure within PLIDA. It does this by linking the datasets supplied by Data Custodians to a central linkage infrastructure called the 'Person Linkage Spine' (**Spine**), so that instead of linking one dataset to another, datasets are linked to the Spine. This creates the capacity for the analytical data in separate source datasets to be linked to one another, via the Spine, in a flexible and efficient manner. Through the Spine, the datasets can be combined as required and extracts of the linked data can be used by Approved Researchers for approved projects. The Spine enables information to be brought together in relation to people who were residents in Australia during a given reference period from 2006 to present.
- 1.7 The ABS is responsible for ensuring that PLIDA data that it makes available to Approved Researchers in the DataLab is provided in a manner that is not likely to enable the identification of an individual (and therefore meets the requirements to be 'de-identified' under the Privacy Act)⁵.
- 1.8 The ABS currently has a range of security arrangements in place for the IT systems used for PLIDA to protect PLIDA data, which:
- 1.8.1 conform with security arrangements set out in the Australian Government Information Security Manual (**ISM**);
 - 1.8.2 ensure that data collection, linkage, and assembly activities for PLIDA datasets are only conducted by a dedicated team in the Secure Data Integration Environment (**SDIE**);
 - 1.8.3 ensure that access to data by approved researchers outside the ABS is conducted only by giving access to analytical data within an ABS environment known as the **DataLab**;
 - 1.8.4 includes a secured internet gateway which is reviewed annually by the Australian Signals Directorate (**ASD**); and
 - 1.8.5 includes an ongoing program of security audits and system accreditations, including the Information Security Registered Assessors Program (IRAP).
- 1.9 Further, PLIDA datasets are handled in accordance with a range of additional privacy protection practices, including:
- 1.9.1 the Five Safes Framework, an internationally recognised approach to managing disclosure risks, which is applied to ensure access to PLIDA data is appropriate. The framework is designed to facilitate safe data release using five elements (Safe People, Safe Projects, Safe Settings, Safe Data and Safe Outputs) which are all assessed independently, but also considered as a whole for each instance of data access;
 - 1.9.2 PLIDA adheres to the 'separation principle' and 'functional separation' in receiving, storing, and curating data for all integration projects. This means that for each dataset, data that contains identifying information about each individual that is the subject of the dataset ('linkage information'), is stored and handled separately from other data about that individual ('analytical information'). Linkage information is stored separately to de-identified, analytical information, and access to these different information sets is restricted so that no individual is able to access both sets of information simultaneously. In addition, functional separation means that ABS staff members undertaking data linkage only have access to the information they need to perform their assigned roles.

⁵ Personal information is de-identified 'if the information is no longer about an identifiable individual or an individual who is reasonably identifiable' (section 6(1) of the Privacy Act).

- 1.9.3 data may be added to PLIDA via a once-off linkage (for a specific research project or projects and will not be retained following the completion of the project(s)), or as a part of an enduring analytical asset that is separate to PLIDA;
- 1.9.4 the ABS is transparent about linkages with PLIDA and approved projects that make use of PLIDA data, through information included on the ABS website.

Previous Privacy Impact Assessments

- 1.10 PLIDA has evolved over time and privacy impact assessments (**PIAs**) have been conducted to assess the privacy impacts associated with PLIDA, with the most recent PIA update occurring in 2022. To date, three PIAs have been conducted on PLIDA (an initial PIA in 2018 (the **2018 PIA**), a PIA update undertaken by the ABS in 2019 (the **2019 PIA Update**) and again in 2022 (the **2022 PIA Update**), with each subsequent PIA process considering planned changes or updates to PLIDA since the previous PIA process.
- 1.11 The ABS has also conducted PIAs on various project-specific PIAs involving PLIDA data during this time, including detailed consideration of the linkages of specific datasets to PLIDA on a case-by-case basis.

2. Overview of the 2024-25 PLIDA Update

- 2.1 The ABS has engaged Maddocks to conduct further PIA update in 2024-25, to:
 - 2.1.1 consolidate consideration of the privacy impacts of some changes or updates to PLIDA since 2022 (taking into account some previous project-specific PIA processes, and other privacy consideration, which has already been undertaken to consider those changes or updates);
 - 2.1.2 consider some future planned updates or changes to PLIDA to ensure continued compliance with the Privacy Act, the Australian Privacy Principles (**APPs**), and privacy best practice; and
 - 2.1.3 facilitate the ABS continuing to take a 'privacy by design' approach to PLIDA.
- 2.2 The 2024-25 PIA Update will cover the following issues (each of which is discussed in more detail in the sections below):
 - 2.2.1 the inclusion of some new or expanded data in PLIDA (in terms of type and volume), where that data will be prepared for linkage in accordance with existing PLIDA processes – specifically:
 - (a) linking new variables drawn from other datasets received from government Data Custodians, which reflect some high-level information about data subjects' interactions with the justice system;
 - (b) linking data received from Data Custodians who are private organisations including 'not for profits'; and
 - (c) consideration of the existing practices around handling of data collected by Data Custodians through surveys.
 - 2.2.2 some expanded outputs from PLIDA data, including:
 - (a) creation of new enduring modules which can be used by Approved Researchers for approved projects, including one which will assist in creating a new Life Course Data Initiative (**LCDI**) dataset;

- (b) use of certain data in PLIDA to supplement or replace data items collected via ABS surveys (including the Census) and undertake more efficient survey operations.
- 2.3 For completeness, the PIA will also consider the potential implementation of some new or changed data handling practices for PLIDA. More information about these additional issues will be provided during stakeholder consultation sessions, but they cover:
- 2.3.1 the potential combining of two internal ABS roles which handle identified data received from Data Custodians (like both of the current roles, the combined role will not involve handling of PLIDA personal identifiers and analytical data at the same time, so that the 'Separation Principle' will continue to apply);
 - 2.3.2 changes to an ABS IT system that is used to prepare PLIDA data;
 - 2.3.3 the mechanism that has been implemented to 'reuse' PLIDA data when Data Custodians have approved its use for the National Disability Data Asset (**NDDA**);
 - 2.3.4 potential access to analytical PLIDA data in the DataLab by Approved Researchers from outside of Australia; and
 - 2.3.5 potential creation of outputs containing aggregated integrated analytical PLIDA datasets for publication on the ABS website, subject to the usual PLIDA and other ABS processes for the release of data.

3. Including new or expanded data in PLIDA

Linking high level administrative crime and justice data to PLIDA

- 3.1 PLIDA already includes many datasets that were originally collected by a Commonwealth or State or Territory government agency as part of their functions and activities, before the agency as Data Custodian provided that data for inclusion in PLIDA (this is often referred to as **administrative data**).
- 3.2 Some analytical datasets for administrative data in PLIDA already contain data from which it is possible to infer that a particular data subject has had an interaction with the justice system in Australia. For example, the data may indicate that a data subject received a social security payment or service that is only available to those leaving or entering prison.
- 3.3 There is a proposal to make some high-level variables in administrative data (received from Australian Government agencies as Data Custodians) available for approved PLIDA projects. This is intended to facilitate research projects designed to enhance informed government policy making, and more efficient and targeted service delivery for persons who have interacted with the justice system (for example, evaluation of employment support services for people who have left prison and are transitioning to work).
- 3.4 The data will use variables such as 'Yes/No' flags (this is sometimes referred to as being only **'high level data'**). The existence of such variables can indicate:
 - 3.4.1 that a data subject is or was in prison (e.g. the variable can be determined because the data shows their Centrelink or another type of payment or service was suspended or cancelled because they are or were in prison; or that they are or were no longer participating in a student or apprenticeship program because they are or were in prison);
 - 3.4.2 that a data subject's partner, or someone else with a relationship to the data subject, is or was in prison (e.g. the variable can be determined from the type of payment or service that the data subject received); and

- 3.4.3 that the data subject, or someone other than the data subject, has had another type of interaction with the criminal justice system (e.g. the variable can be determined because the data subject received crisis or other payments or services related to family or domestic violence).
- 3.5 The data (including the new variables) will not contain more detailed information about the interaction with the criminal justice system – e.g. it will not show the reasons why a data subject was in prison, what crime was committed (or alleged), or any other details about their interaction with the police, courts, or family and domestic violence protection or support systems.
- 3.6 Existing PLIDA processes mean that the ABS will ensure:
- 3.6.1 relevant Data Custodians must confirm that there is legislative or other authority for the inclusion of the new variables in PLIDA, and their use for approved projects;
- 3.6.2 that the new variables will only be included in datasets that do not include any direct identifiers about the data subject (that is, they will be included in analytical datasets only, and not in the Spine used to link different datasets); and
- 3.6.3 the new variables will only be able to be used for projects that have been approved under the data governance framework for PLIDA;
- 3.6.4 outputs from those projects will be subject to existing PLIDA processes, designed to ensure that no individual can be identified from those outputs; and
- 3.6.5 all of the usual security protections for PLIDA data will apply to those variables (including technical IT protections and data governance measures, including to restrict and monitor access).
- 3.7 Any proposal to include more detailed information in PLIDA about a data subject’s interaction with criminal or justice systems (beyond that contained in the new high-level variables), or any proposal to include a dataset in PLIDA that contains (or is compiled from) police records, court or corrective services records, or any other data about a data subject’s criminal history, will continue to require the undertaking of a separate PIA process. This means that any proposal to include State and Territory police, courts and prison data within PLIDA would be the subject of a separate PIA process). For example, the ABS is currently undertaking a PIA process for compilation of a new Crime Justice Data Asset (**CJDA**) which will link certain State and Territory police, courts and prison data.⁶

Questions for stakeholders:

- What are your thoughts on the use of high level variables relating to crime and justice in PLIDA projects?
- Should there be any additional restrictions placed on who can be authorised to access these new types of data?
- Are there any types of research projects for which use of these new variables will always be inappropriate?

⁶ As the privacy impacts will be considered as part of project-specific PIA processes, the 2024-25 PLIDA Update PIA will not examine any of these projects in any detail.

Linking not-for-profit (NFP) organisational data to PLIDA

- 3.8 PLIDA already includes some private sector⁷ data that is received from organisations who operate in a 'not for profit' manner, that is, they do not operate for the profit, personal gain or other benefit of particular people (**NFP organisations**).⁸ Many of these NFP organisations are funded by government(s) to conduct their operations and to collect particular data.
- 3.9 Some data originally collected by NFP organisations is subsequently provided to a government agency, and then provided to PLIDA by that Data Custodian as administrative data. Other data is already provided to PLIDA directly by an NFP organisation as the Data Custodian. For example, in 2023-24 the ABS undertook an Expanded Health Data Linkage PIA, which considered the inclusion of cancer registry data provided by State and Territory NFP organisations operating those registers, for projects to help understand patient journeys through the health care systems.
- 3.10 The 2024-25 PLIDA PIA Update will build upon this consideration to assess the inclusion of other types of data from NFP organisations beyond health data.
- 3.11 It is only intended that data would be collected from large NFP organisations in relation to services that are funded by government, where there would be significant benefit and utility to include the data in PLIDA.⁹ Such data might include that from:
- 3.11.1 large childcare providers, for projects to help understand pathways through childcare facilities (including to understand impacts or the extent of any disadvantage children may have); or
 - 3.11.2 organisations who deliver services to people participating in their programs funded by government (for example, programs that provide supports to people experiencing homelessness or disadvantage by providing accommodation or housing solutions), for research projects to help understand the outcomes and pathways of participants in those programs, and/or to evaluate the effectiveness of those programs).

Questions for stakeholders:

- What are your thoughts about data being received from NFP organisations (compared to other types of Data Custodians)?
- Should this data be treated any differently to other types of data that are collected from other types of Data Custodians?
- Are there any types of research projects for which use of data received from NFP organisations will always be inappropriate?

⁷ Private sector data was broadly covered in the 2022 MADIP PIA Update. The ABS undertakes additional transparency measures when considering new private sector data linkages to PLIDA, including NFP data.

⁸ This definition of NFP is sourced from the Australian Charities and Not-For-Profits Commission, available here: <https://www.acnc.gov.au/tools/topic-guides/not-for-profit#:~:text=A%20not%2Dfor%2Dprofit%20is,%2Dfor%2Dprofits%20are%20charities.>

⁹ It is not intended that datasets would be collected from any NFP organisations who are a 'small businesses operator' as defined in the Privacy Act (generally, an entity will be a small business operator if it has an annual turnover of \$3m or less for a financial year, although there are some exceptions to this rule).

‘Sensitive’ and ‘non-sensitive’ survey data

- 3.12 From time to time, a Data Custodian (including the ABS in its role as a Data Custodian) may propose that data that was obtained through a survey of data subjects (**survey data**) should be included as a dataset within PLIDA.
- 3.13 Such survey data may contain:
- 3.13.1 information that falls within one of the categories set out in the definition of ‘sensitive information’ under the Privacy Act, such as personal information about an identified individual’s racial or ethnic origin, sexual orientation, criminal record or health, or their genetic or biometric information (**sensitive survey data**); or
 - 3.13.2 information that does not meet that legislative definition, but nevertheless has an inherent degree of sensitivity (for example, information about an individual’s experience with family, domestic or sexual violence) (**inherently sensitive survey data**); or
 - 3.13.3 other personal information that does not fall into either of the above categories (**non-sensitive survey data**). Inclusion of non-sensitive survey data into PLIDA was considered as part of the 2019 PIA Update.
- 3.14 Currently, if survey data includes sensitive survey data, or inherently sensitive survey data, a separate PIA process is undertaken before that survey data is included in PLIDA and prepared for integration. For example, separate PIA processes have been undertaken to link data from the National Health Survey (the PIA was published in August 2018), the Survey of Disability, Ageing and Carers (which was covered in the 2019 PIA Update), and the National Study of Mental Health and Wellbeing (a PIA was published in November 2020).
- 3.15 Through consultation, the ABS seeks to understand if this continues to be the appropriate approach to the inclusion of survey data in PLIDA.

Questions for stakeholders:

- Do you agree with the current approach (i.e. a specific PIA should be undertaken if the survey data to be included in PLIDA contains ‘sensitive information’ or ‘inherently sensitive’ information)?
- How should ‘inherently sensitive’ survey data be defined?
- Are there any types of inherently sensitive survey data that should never be linked?

4. Creation of expanded PLIDA outputs

Creation of new PLIDA modules

- 4.1 The ABS has an established process for creating **PLIDA modules** from PLIDA data. Modules consist of ‘pre-linked’ datasets which can be made accessible to Approved Researchers for approved projects. Modules are ‘enduring’ datasets within PLIDA (they are not created specifically for any particular approved project), which are regularly updated and available for use in any approved research project. All Modules are approved by source data custodians.
- 4.2 Modules are accessible by Approved Researchers in the DataLab via the PLIDA Modular Product (**PMP**), which includes the deidentified data items and reference periods for each module available to Approved Researchers.

- 4.3 There is a significant demand for additional modules to be created using existing and new PLIDA datasets. These include the potential new modules described below.

New Core Modules

- 4.4 The ABS has created five new PLIDA modules (**Core Modules**) using existing PLIDA data, some of which have replaced existing modules in the PMP. The new Core Modules are:

4.4.1 the **Core Demographics Module**, which contains demographic items derived from data across the PLIDA datasets such as data subjects' age, gender, death dates, country of birth and highest level of education attained;

4.4.2 the **Core Indigenous Module**, which contains information about whether or not a data subject has identified as Aboriginal and/or Torres Strait Islander.¹⁰ This module replaced the 'ever identified' derived Indigenous status currently contained within the previous 'Combined Demographics – Indigenous Module'. The new module continues to use data drawn from existing administrative data within PLIDA.

This module also includes code or pseudo-code for derivation algorithms, to allow Approved Researchers to construct their own Aboriginal and/or Torres Strait Islander status indicators appropriate to their research needs – it is anticipated that this module will support Approved Researchers in navigating the current complexities and challenges in deriving Aboriginal and/or Torres Strait Islander status from administrative data in PLIDA.

Creation of the new module was supported by Aboriginal and/or Torres Strait Islander stakeholders separately consulted by the ABS about this change.

4.4.3 the **Core Locations Module**, which contains existing geographic information about data subjects within PLIDA using information from administrative data within PLIDA. It includes data items for data subjects such as their State or Territory of residence, an address register ID (**ARID**) with the start and end dates of residence at the ARID, and information about the data subject's dwelling (structure, type of dwelling, and derived residential or commercial use);

4.4.4 the **Core Scoping Module**, which collates information in PLIDA about data subjects' residence and activity in Australia. This module contains information about data subjects' 'vitals' (i.e. the month and year of birth and death), flags that indicate physical presence or residential status in Australia, and flags that indicate whether a data subject is in particular administrative data sets in PLIDA. It is intended that this module will better support Approved Researchers undertaking projects in population scoping.

4.4.5 the **Core Relationships Module**, which will support research projects examining family and household analysis. It contains information about data subjects' family relationships in Australia drawn from other PLIDA datasets (such as their relationship status, start/end dates of relationship records, and relationship codes such as for 'partner' or 'parent'). This module will be used to assist with the Life Course Data Initiative (**LCDI**) project (discussed further below).

- 4.5 The Spine is used to assemble the data, primarily from existing administrative data within PLIDA. However, to create the Modules, the ABS also needed to include some other datasets in PLIDA (using the existing PLIDA data governance framework) which were previously collected by the ABS but not previously included within PLIDA.

¹⁰ We have used the phrase 'Aboriginal and/or Torres Strait Islander' to describe persons who identify as an Indigenous Australian, to respect the term that we understand most Aboriginal and/or Torres Strait Islander people prefer to use, but noting that in Australia there are many Indigenous nations, languages and cultures.

Life Course Data Initiative (LCDI)

- 4.6 The LCDI is a pilot funded under the Targeting Entrenched Disadvantage Package, which will run from FY2023-24 to FY2026-27. One aim of the LCDI is to create a new dataset (to be known as the **Life Course Dataset**), which can be used for Approved Researchers in selected research projects to inform long-term policy responses to disadvantage (by generating useful and actionable insights to help communities and policymakers understand and address child disadvantage).
- 4.7 The ABS will establish the Life Course Dataset, which will expand the existing PLIDA by linking additional datasets that may include data from Commonwealth, State and Territory governments and NFP organisations. Some of the new datasets included in the Life Course Dataset will be provided by the Australian Capital Territory (ACT) and South Australian governments.
- 4.8 The Life Course Dataset will contain analytical content drawn from existing PLIDA datasets (including the Core Modules), and five new datasets that will be included specifically for the LCDI project. To improve data coverage over the early years of Australians' lives, particularly from ages 0-14,¹¹ the Life Course Dataset will include analytical information across many aspects of people's lives including health, education, employment, security, and housing.
- 4.9 The pilot LCDI will begin with a small number of research projects: led by the jurisdictional partners, the ABS, and other select research organisations. The ABS will use the Life Course Dataset in various ways to better understand disadvantage. To support this, the ABS will use the Life Course Dataset to develop household and family structures, which can be used by Approved Researchers in their analyses on topics relating to disadvantage, particularly child disadvantage.
- 4.10 More detail about these structures is set out below:
- 4.10.1 **Household structures:** these will identify all data subjects in the Life Course Dataset who live together in the same household at a particular time, and how these persons change over time. The ABS will do this by:
- (a) using Core Locations Module data that indicates that data subjects cohabit (e.g. if data subjects have the same address register ID number (**ARID**) at a point in time, they will be part of the same household structure at that time); and
 - (b) where this is not possible (e.g. because specific address data for a data subject is not available) endeavouring to use other data in the Life Course Dataset (or data which can be added to the Life Course Dataset) to infer or input that data subjects are likely to have cohabited at a particular time.¹²
- 4.10.2 **Family structures:** these will identify, for each dependent child in the Life Course Dataset (i.e. a child aged 14 or under), data subjects who have a relationship with that child (e.g. because they are the child's parent or guardian, a partner of a child's parent or guardian¹³, sibling, grandparent or other family member). At times those other data subjects may be cohabiting with the child, but other times they may not be cohabiting. The ABS will do this by:
- (a) using the Core Relationships Module and other data in the Life Course Dataset that explicitly indicates a relationship between a child and other data subjects; and

¹¹ More information here: <https://www.abs.gov.au/about/key-priorities/life-course-data-initiative>.

¹² For example, other data may show that both parents of a young child have moved to a new address, so it might be inferred that it is likely that the child will also have moved to that address.

¹³ The data will also show if a data subject stops having that relationship with a child.

- (b) where relationship information for a child is missing, endeavouring to infer or estimate relationships using regression analysis or machine learning models.

4.11 The ABS will undertake data validation activities, to check the quality of the household and family structures identified. This may involve using other datasets to see whether similar proportions of relationships are found using other sources (e.g. using released Census data to see whether similar proportions of single parent families have been identified), and examining the proportions of relationships that were created by direct, inferred or estimated methods. These activities will be used to produce documentation for Approved Researchers that describes the methods, datasets and potential quality limitations of the structures.

Questions for stakeholders:

- What are your thoughts about the creation of the new Core Modules?
- Are the existing PLIDA arrangements appropriate for access to the new Core Modules?
- Do you have any thoughts about the creation of either the household or family structures for the LCDI?

Expanding the use of PLIDA to support statistical production (including the Census and ABS Household Surveys)

4.12 From time to time, proposals have been considered, on a case by case basis, to use PLIDA data to support statistical production of particular data.

4.13 For example, separate PIA processes have been, or are currently being, considered for the use of PLIDA data:

4.13.1 to improve the efficiency of collection and quality of data collected by the ABS as part of the Census (e.g. to use information that indicates whether a property was not occupied on Census night, so that time and resources are not wasted in following up occupants to complete a Census form, thereby making the collection process more efficient); or

4.13.2 to improve data quality and save survey respondents time and effort in completing surveys, by supplementing survey data with data items already included in PLIDA (e.g. to use information about PBS medications dispensed to a person to supplement their answers to the NHS survey questions; to use PLIDA data to fill in missing information in the Census record or to add additional variables for questions that were not asked).¹⁴

4.14 The ABS wishes to consider more broadly the information flows and privacy risks associated with some key use cases, for which regular requests for PLIDA data to be used to support statistical production of ABS surveys (including the Census), specifically the use of PLIDA data for:

4.14.1 data supplementation – the use of PLIDA administrative data (i.e. data received from a government agency as the Data Custodian) to add more data items to other data which is collected as part of an ABS survey;

4.14.2 data substitution – the replacement of survey data with administrative data contained in PLIDA (e.g. information in the survey about the data subject’s healthcare, income or tax, or childcare costs can be replaced with data in the PLIDA data which is considered more reliable); and

¹⁴ As these issues were, or are being, considered through separate PIA processes, they are not within the scope of the 2024-25 PIA Update.

Attachment 3 Copy of material provided to stakeholders – Consultation slides



Please *mute* your
telephone/mobile phone or computer





Privacy consultation for the 2024–25 Person Level Integrated Data Asset (PLIDA) PIA Update

October 2024

AUSTRALIAN BUREAU OF STATISTICS
Informing Australia's important decisions





Our Story.
Our Future.

The Australian Bureau of Statistics acknowledges the Traditional Owners and Custodians of Country throughout Australia and recognises their continuing connection to land, waters and community. We pay our respects to their cultures and Elders, past and present.



How will the session work?



Mute your computer microphone or telephone when not actively speaking



Raise your hand, chat or email: privacy@abs.gov.au



Email written feedback within two weeks from this session



4



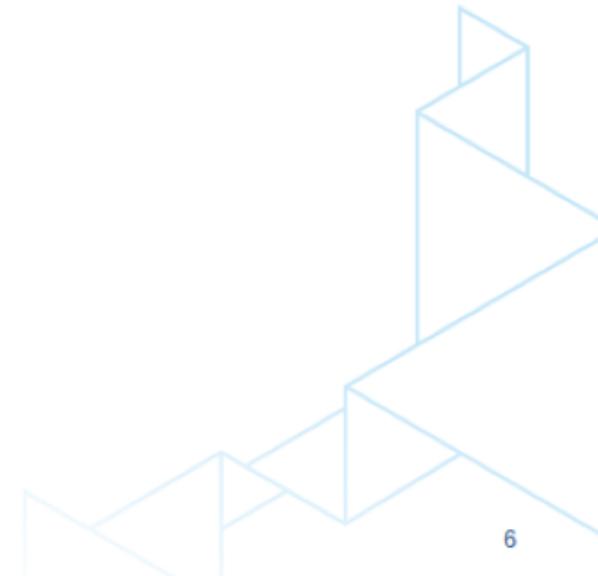
Outline of session



PART 1	Privacy Impact Assessments and consultations
PART 2	Overview 2024–25 PLIDA PIA Update
PART 3	Data integration overview and keeping data safe
PART 4	Privacy considerations

What is privacy?

- ▶ Privacy is a human right that underpins freedom of association, thought and expression, as well as freedom from discrimination.
- ▶ Today we focus on information privacy. This type of privacy is about protecting our information, including:
 - Who we are
 - What we do
 - What we believe



How do we assess privacy?



- ▶ A **privacy impact assessment** (PIA) identifies the impacts that a project might have on the privacy of individuals, and provides strategies for managing, minimising or eliminating those impacts.
- ▶ The ABS undertakes PIAs in line with the [Australian Government Agencies Privacy Code](#) and guidance from the Office of the Australian Information Commissioner
- ▶ Privacy experts at **Maddocks** are leading the 2024–25 PLIDA PIA Update
- ▶ The final PIA report and a response from the ABS will be published on the [ABS website](#) in early/mid 2025



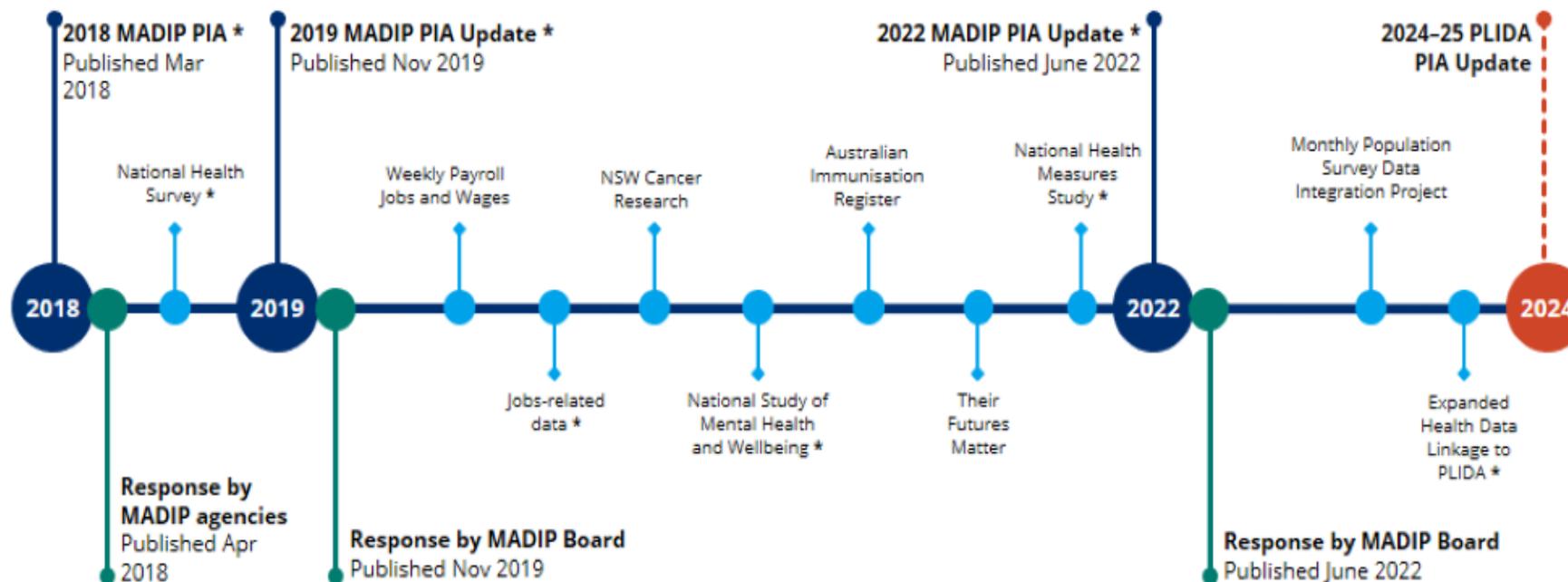
Your views about privacy are important



We want to know ...

- What privacy issues are important to you?
- Are we addressing them? What is missing?
- How can we manage privacy better?

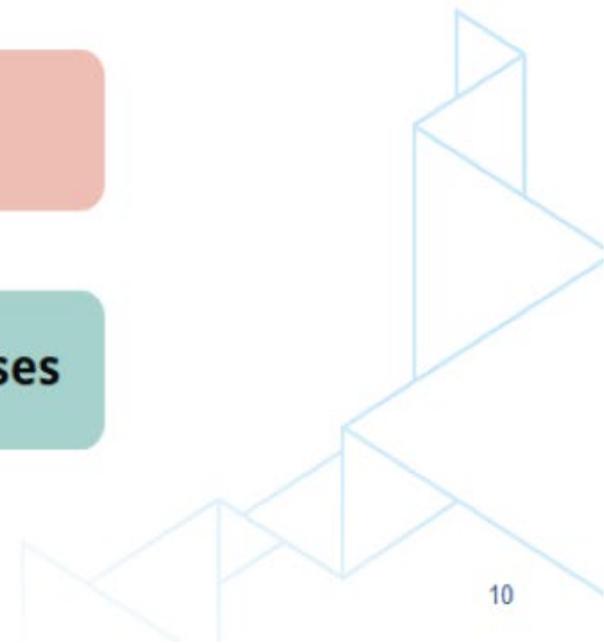
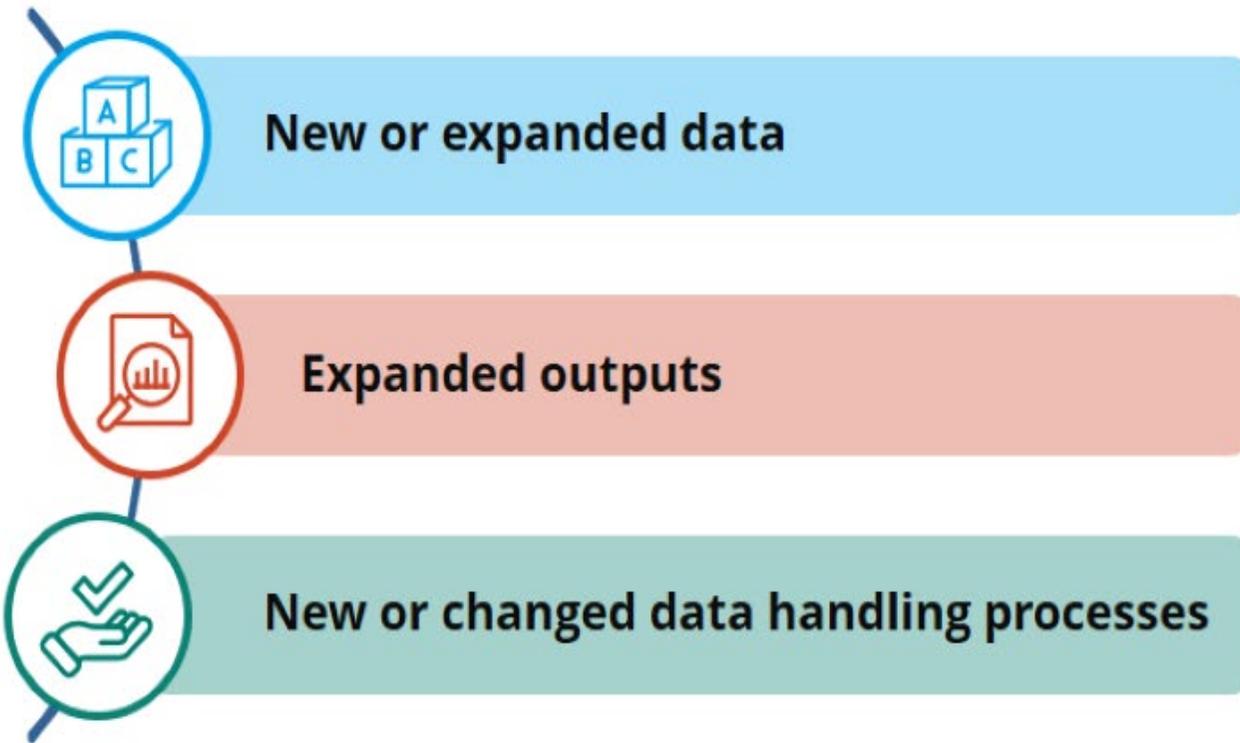
Timeline for MADIP/PLIDA and related PIAs



All documentation available on the [ABS Privacy Impact Assessments](#) webpage

- Project-specific PIAs involving PLIDA data
- ✱ Consultation undertaken
- - - Currently undertaken

High level changes for the PIA Update



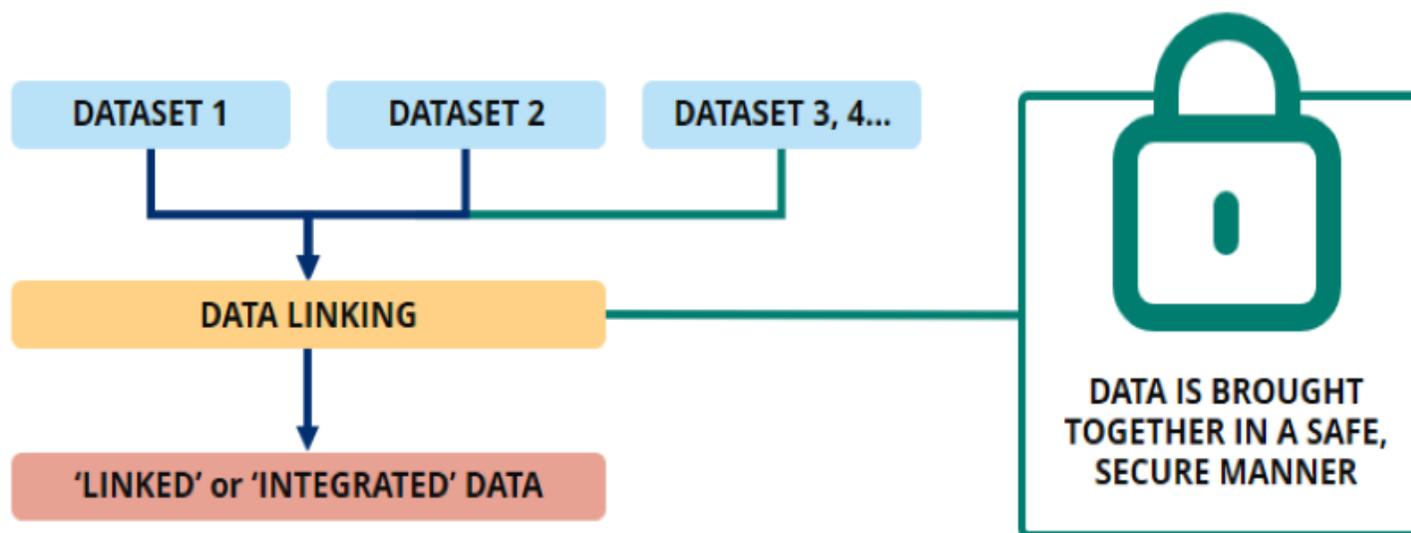
Questions and discussion

- ▶ In this section, we have covered:
 - PIAs and why we are consulting
 - Timeline of MADIP/PLIDA and related PIAs
 - High level changes considered in the PIA Update



Data integration

- ▶ Combining two or more sets of information about people
- ▶ Information without personal identifiers (such as names and addresses) will be used for data analysis



Overview of PLIDA



- ▶ PLIDA is a partnership among Australian Government agencies that has developed a secure and enduring approach for combining nationally important datasets to create a comprehensive picture of Australia over time.
- ▶ PLIDA data is securely held by the ABS – access to de-identified data is only made available to approved researchers for approved purposes in a secure analytical environment.

PLIDA Uses

- Answer policy questions
- Program evaluation
- Empirical research on socio-economic issues
- **No identification of individuals**
- **Used only for statistical and research purposes (never for compliance)**

PLIDA Partner Agencies

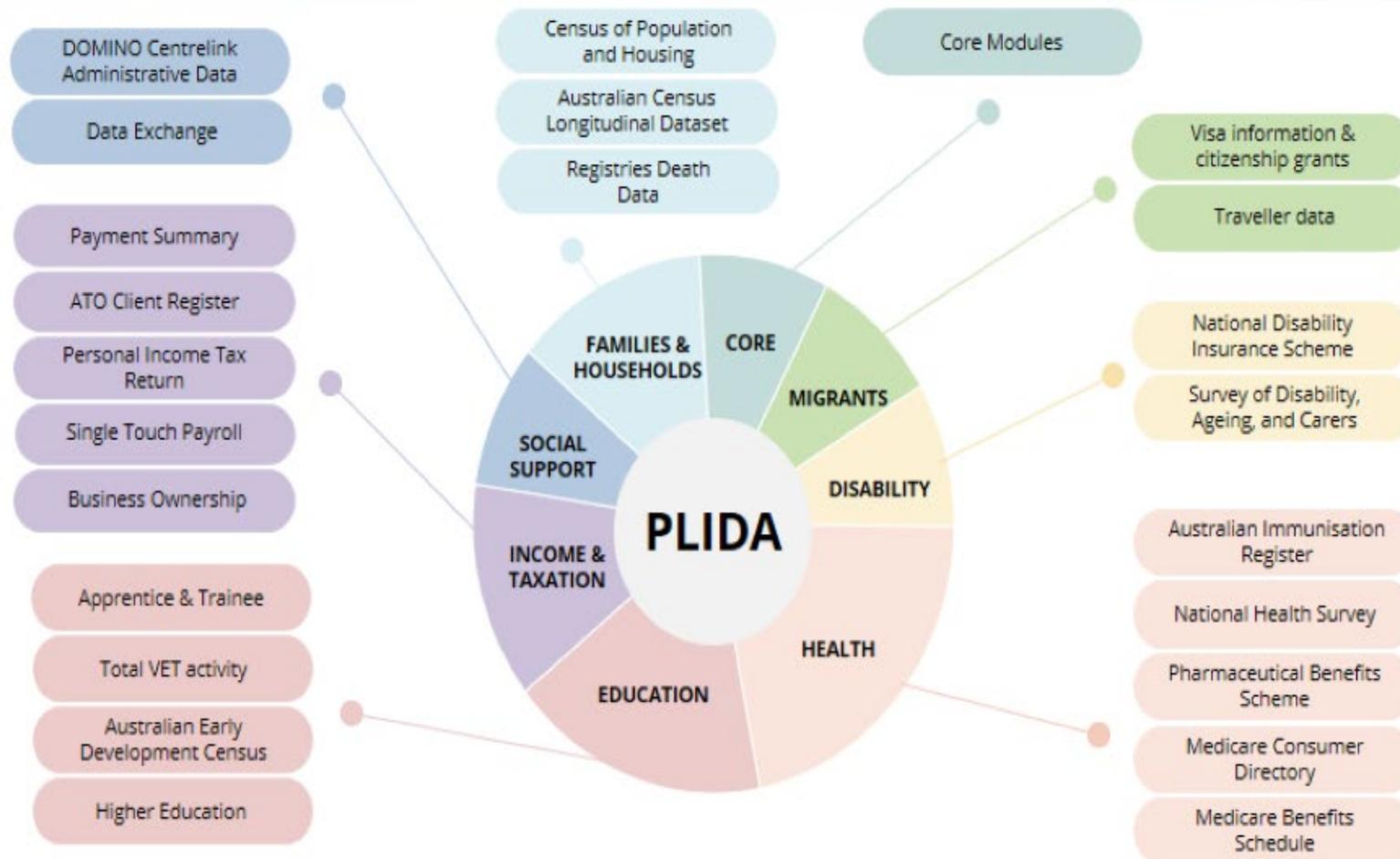
- Australian Bureau of Statistics
- Australian Taxation Office
- Department of Health and Aged Care
- Department of Social Services
- Department of Education
- Services Australia
- Department of Home Affairs

Nationally important datasets



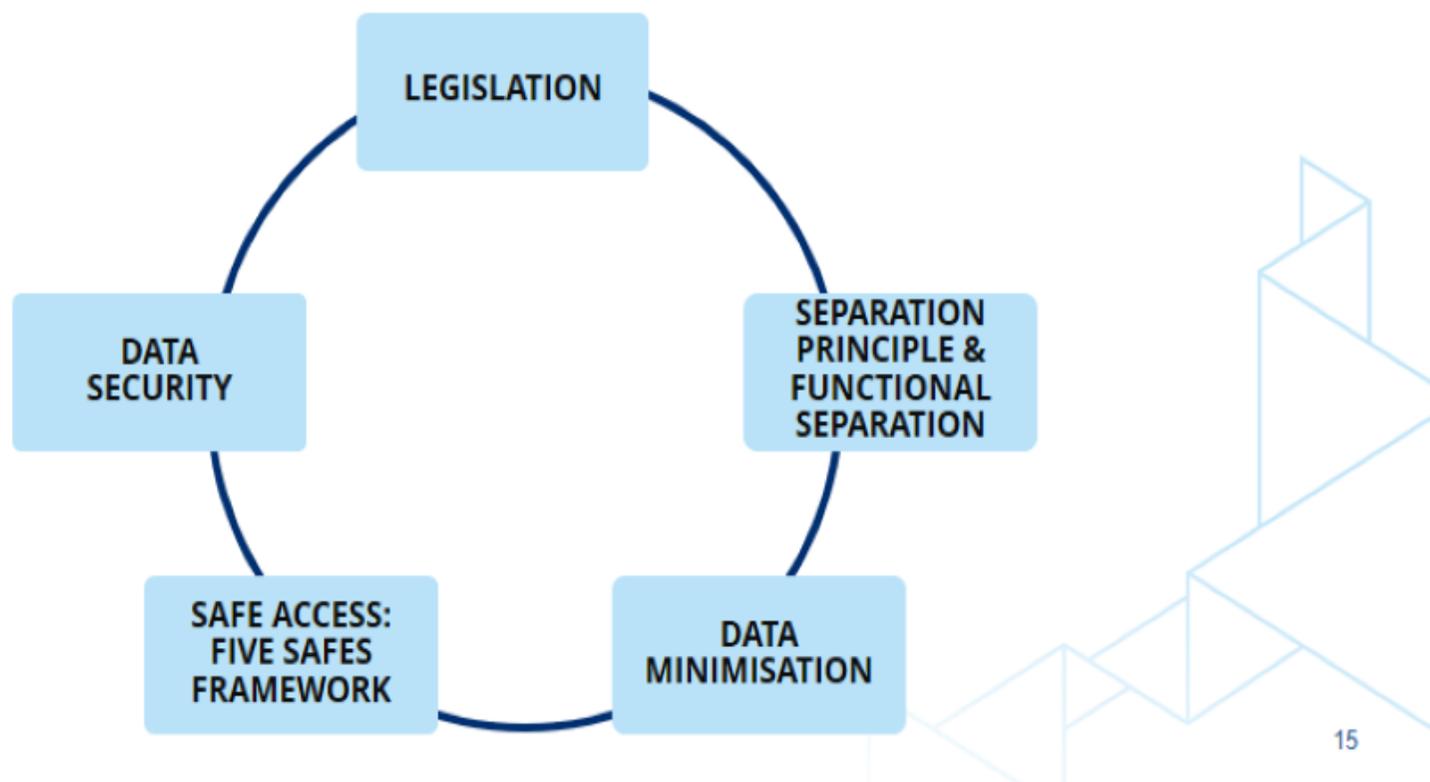
Visit the [PLIDA Research Projects](#) for the full list of projects

PLIDA Datasets by domains



Keeping data safe

- ▶ There are many protections working together to keep PLIDA data safe:



Legislation and standards



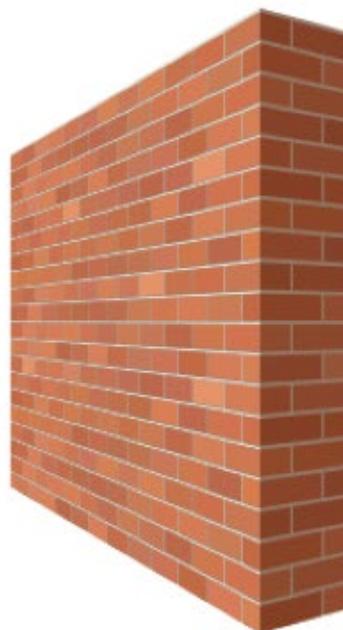
- ▶ **Census and Statistics Act 1905** – ABS collects PLIDA data under this Act
- ▶ **The Privacy Act 1988** – provides protections for use of personal information and allows data use, where authorised by law
- ▶ **Data custodian's legislation** – each data custodian shares data with the ABS for PLIDA lawfully which may come from the data custodians:
 - Establishing legislation
 - Legislation that covers the data
 - Specific data sharing legislation
 - Other legislation
- ▶ **Australian Government standards** – provides strong security arrangements for systems and access

Do you have any questions or concerns about the legal pathways for sharing and use of PLIDA data?

Separation of information

Personal identifiers such as:

- Name
- Address



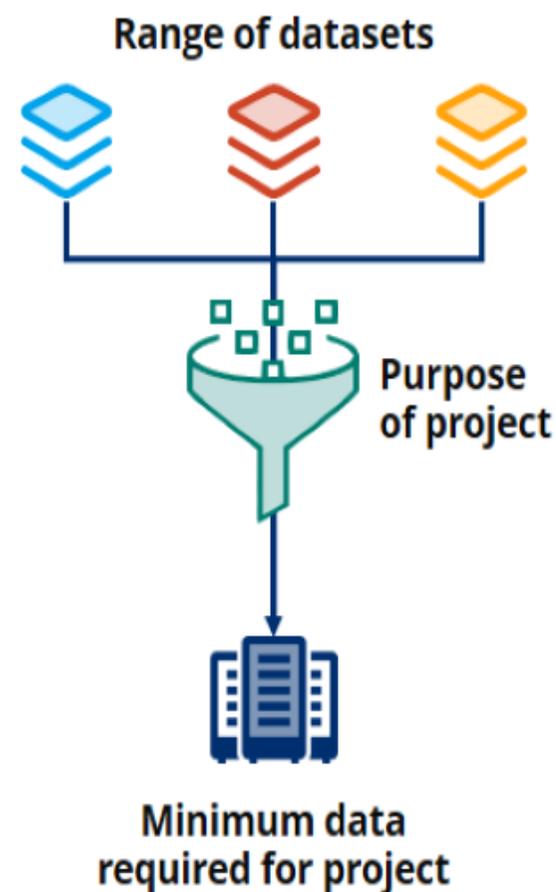
Analytical information such as:

- Income
- Health services use
- Types of Government payments

No person can see all of your information together at any point

Data minimisation

- ▶ Only data necessary for the purposes of the project is shared and used
- ▶ Data is only accessed on a 'need to know' basis
- ▶ Data is only brought together for particular projects
- ▶ Projects must justify requests for sensitive information



Safe access – Five Safes Framework



 Safe People	Is the researcher authorised to access and use the data appropriately?
 Safe projects	Is the data to be used for an appropriate purpose?
 Safe settings	Does the access environment prevent unauthorised use?
 Safe data	Has appropriate and sufficient protection been applied to the data?
 Safe output	Are the statistical results non-disclosive?

Data Security



Regular independent reviews of security arrangements and an ongoing program of security audits (IRAP)

IT security arrangements that conform with government standards for information security (the **Australian Government Information Security Manual**)

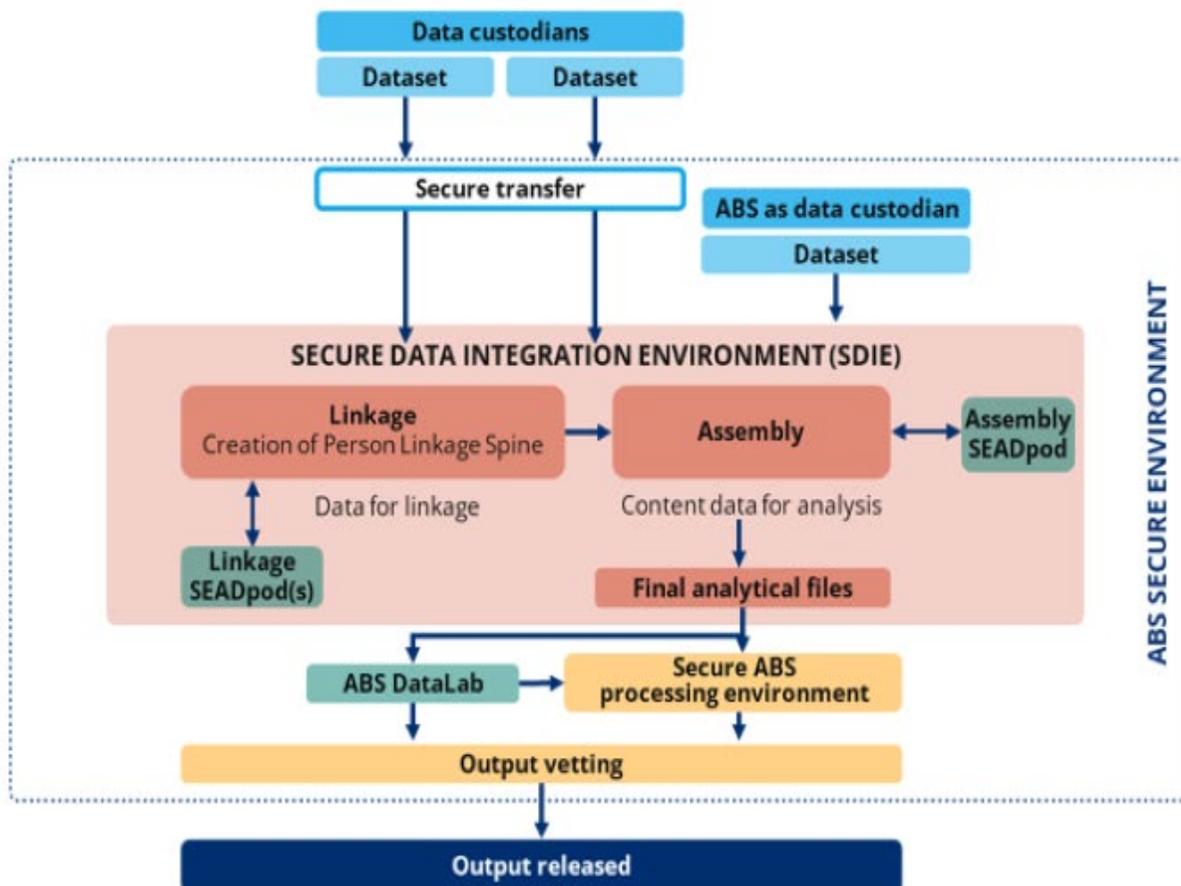
Clear accountabilities and **risk management** processes

Logging and monitoring of access and use of information

Staff **security checks** and **restricted access** to data

Secured internet gateway reviewed annually by the Australian Signals Directorate

PLIDA conceptual data flow



Information is:

- ▶ provided into a secure system
- ▶ separated and prepared for analysis
- ▶ analysed in ABS analytical/processing environments
- ▶ only allowed to exit the system in de-identified manner for approved projects and users

Questions and discussion

- ▶ Keeping data safe
- ▶ Overview of data integration
 - Legislation and standards
 - Separation Principle
 - Data Minimisation
 - Five Safes Framework
 - Data Security
- ▶ PLIDA conceptual data flow



New or expanded data



Adding new or expanded data ensures PLIDA remains relevant and of a high quality to inform Australia's evolving data needs:

Linkage of high-level administrative crime and justice data - case study 1	Linkage of not-for-profit data - case study 2	Linkage of survey data (with sensitive data items)
Linking administrative crime and justice data with PLIDA.	Linking a broad range of not-for-profit organisation data.	Broader inclusion of survey data with sensitive data items in PLIDA

High level administrative crime and justice data

– Case study 1



- ▶ **Proposal:** Link high-level administrative crime and justice related data to PLIDA
- ▶ **Public benefit:** Inform evidence-based policy making and more efficient service delivery for persons who have interacted with the justice system.
- ▶ **Types of data and sources:** Data obtained from routinely collected administrative government data. This data covers whether a person has been in contact with the justice system to assess eligibility and receive support from government services.
 - Examples include:
 1. Government benefits (such as revocation or restoration of Disability Support Pension when an individual goes to or is released from prison)
 2. Incentive payments
 3. Crisis payments
- ▶ **Not in scope:** Detailed crime and justice data relating to individuals' interactions with the justice system such as details of offences, police or courts data etc.

Linking Not-for-Profit (NFP) organisation data – Case study 2



- ▶ **Proposal:** Link a broader range of not-for-profit sector data to PLIDA on a case-by-case basis only.
- ▶ **Public benefit:**
 - Delivering services to the vulnerable population most in need
 - Supporting the development of a learning system based on research evidence and effectiveness of policies and services.
- ▶ **Types of data and sources:** Data collected from large NFP organisations in relation to services that are funded by government.
 - For example:
 1. A not-for-profit organisation which aims to provide children the best possible start in life through access and quality early education, especially those in remote communities or with additional needs.
 2. This data could provide information on developmental vulnerability amongst pre-school children, especially in remote communities and children with additional needs.

New or expanded data



Linkage of high-level administrative crime and justice data - case study 1	Linkage of not-for-profit data - case study 2	Linkage of survey data (with sensitive data items)
Linking administrative crime and justice data with PLIDA.	Linking a broad range of not-for-profit organisation data.	Broader inclusion of survey data with sensitive data items in PLIDA

Questions for discussion:

- 1 What are your thoughts on the use the new data or expanded data in PLIDA projects?
- 2 Should there be any additional restrictions placed on who can be authorised to access these new or expanded types of data?
- 3 Are there any types of research projects for which use of the new or expanded data will always be inappropriate?

Expanded outputs



Adding new or expanded data ensures PLIDA remains relevant and of a high quality to inform Australia's evolving data needs:

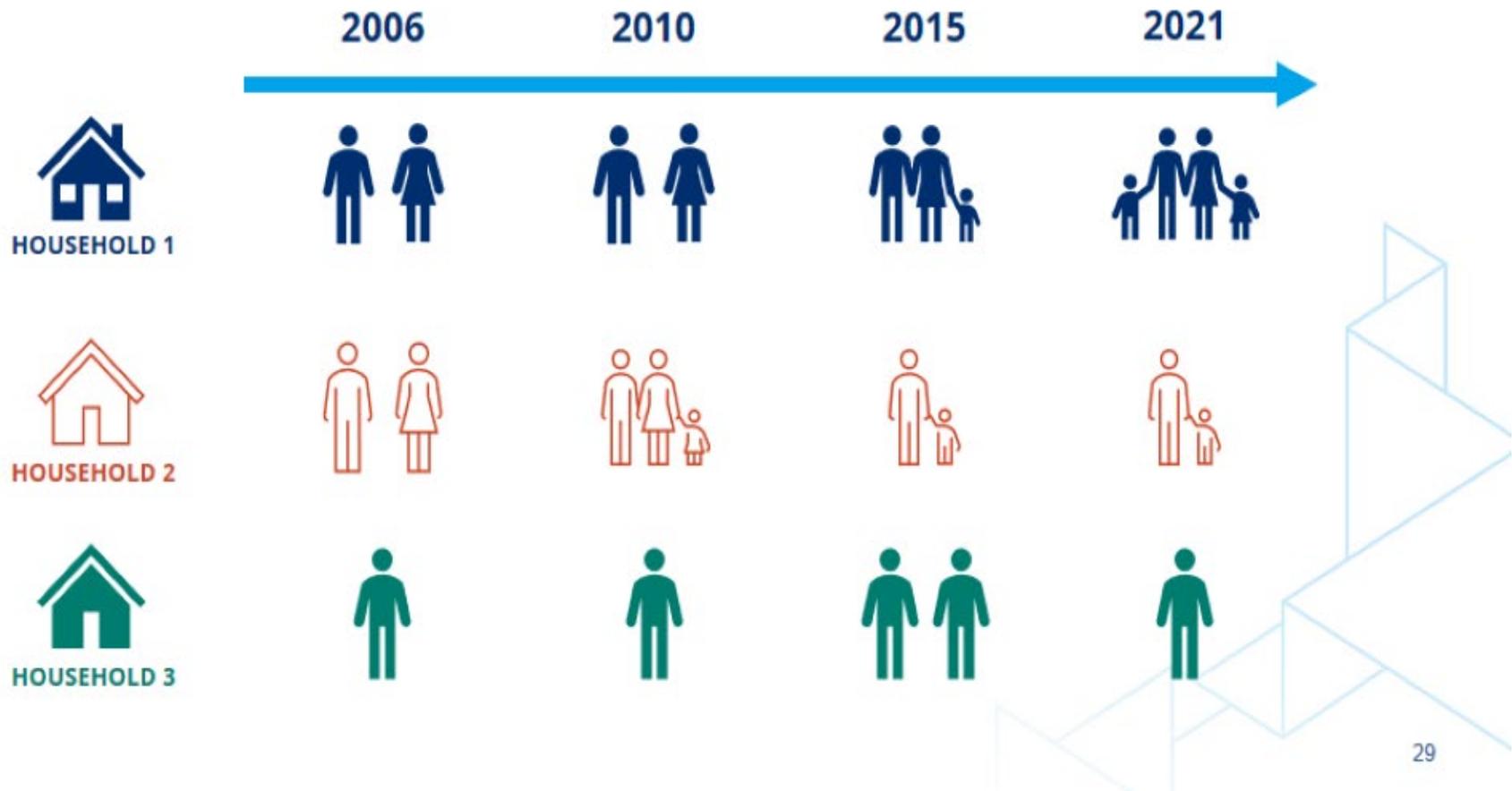
Creation of new PLIDA datasets - Case study 3	Expanded use of PLIDA to support statistical production
<p>Creation of new datasets/modules available to researchers in the ABS DataLab on an enduring or re-usable basis.</p> <p>These include: Core Demographics, Core Indigenous, Core Locations, Core Scoping, and Core relationships Modules.</p>	<p>Standard data flows for using PLIDA data for supplementation, substitution, and editing ABS survey and Census data, and using PLIDA data to develop a dwelling propensities asset to support survey collection.</p>

Life Course Dataset – Case study 3

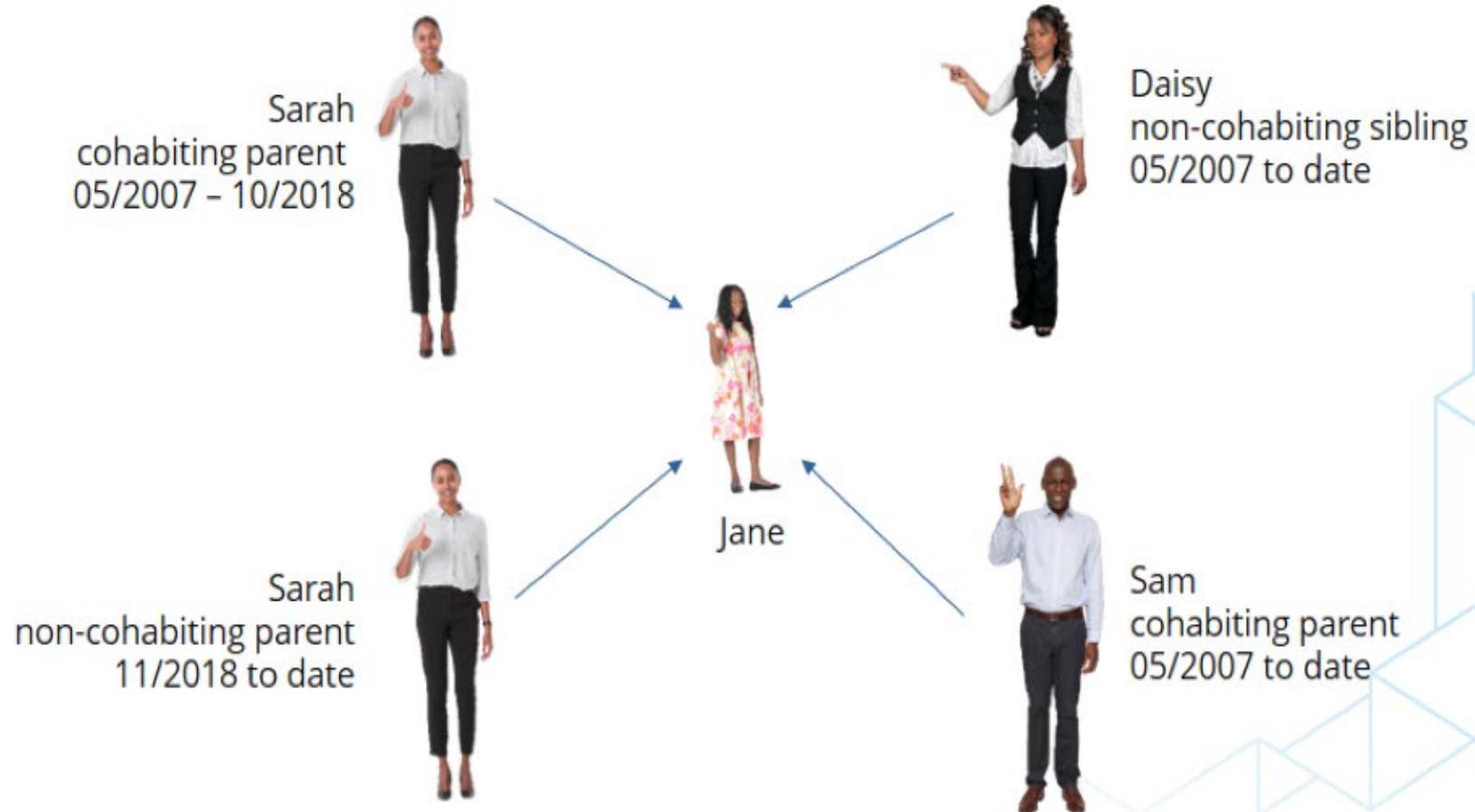


- ▶ **Proposal:** Create the Life Course Dataset by linking existing PLIDA data to administrative data sources.
 - Focuses on improving data coverage for over the early years of Australians' lives, particularly ages 0-14
 - Used by select researchers to understand issues like child disadvantage, to help inform to help understanding and provide new insights into disadvantage in Australia
- ▶ **Public benefit:** The LCDI is expected to help:
 - Researchers understand who and what communities are experiencing disadvantage and what factors may be contributing to those experiences; and
 - Inform public policy response to ease individual or location-based disadvantage.
- ▶ **Types of data and sources:**
 - The ABS will use the Life Course Dataset to conduct its own analyses related to disadvantage. Using PLIDA data, the ABS will also develop:
 1. Household structures; and
 2. Family structures.

LCDI – Household Structures



LCDI – (child-centred) Family Structures

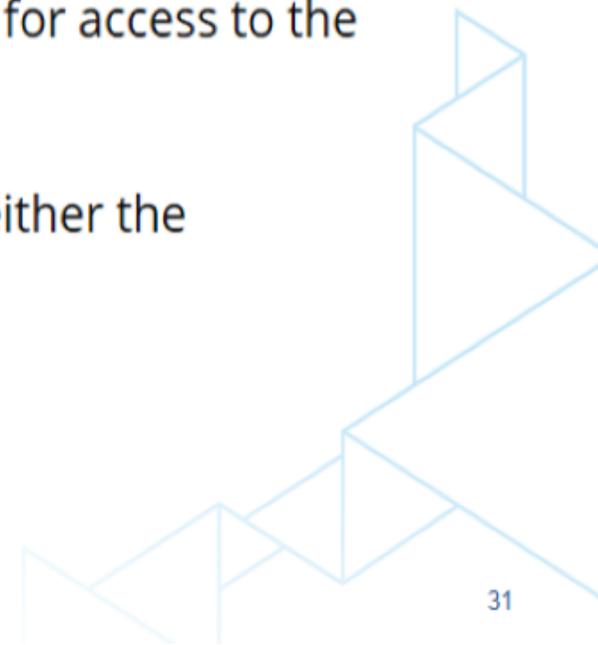


Expanded outputs



Questions for discussion:

- 4 What are your thoughts on the creation of the new Core Modules?
- 5 Are the existing PLIDA arrangements appropriate for access to the new Core Modules?
- 6 Do you have any thoughts about the creation of either the household or family structures for the LCDI?



31

New or changed data handling practices



The PIA will consider the privacy considerations for the following potential changes in PLIDA data handling practices:

Changing implementation of functional separation - case study 4	Re-use of PLIDA data for the NDDA and ANDII - case study 5	Migrating the SDIE into the ANDII ICT environment - case study 5	International researcher access to PLIDA data in DataLab
Combining librarian and linker roles	Re-use of existing and future PLIDA data modules for NDDA/ANDII using the DATA Scheme subject to data custodian approval	Document potential future change in information flows and data environments for use of PLIDA by ABS	Consideration of providing international researchers remote access to integrated microdata (covered in 2022 MADIP PIA Update)

Changing implementation of functional separation

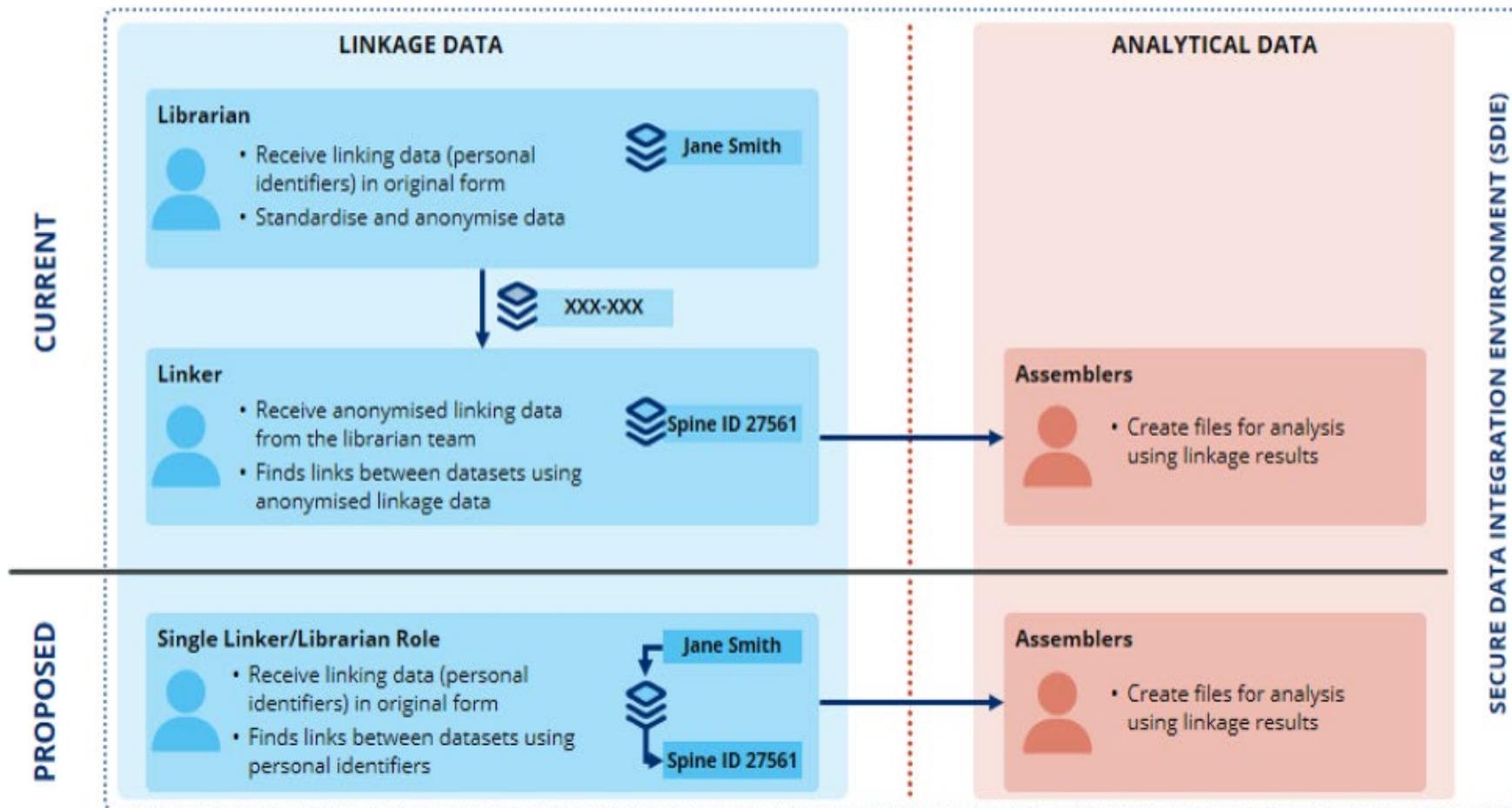
– Case study 4



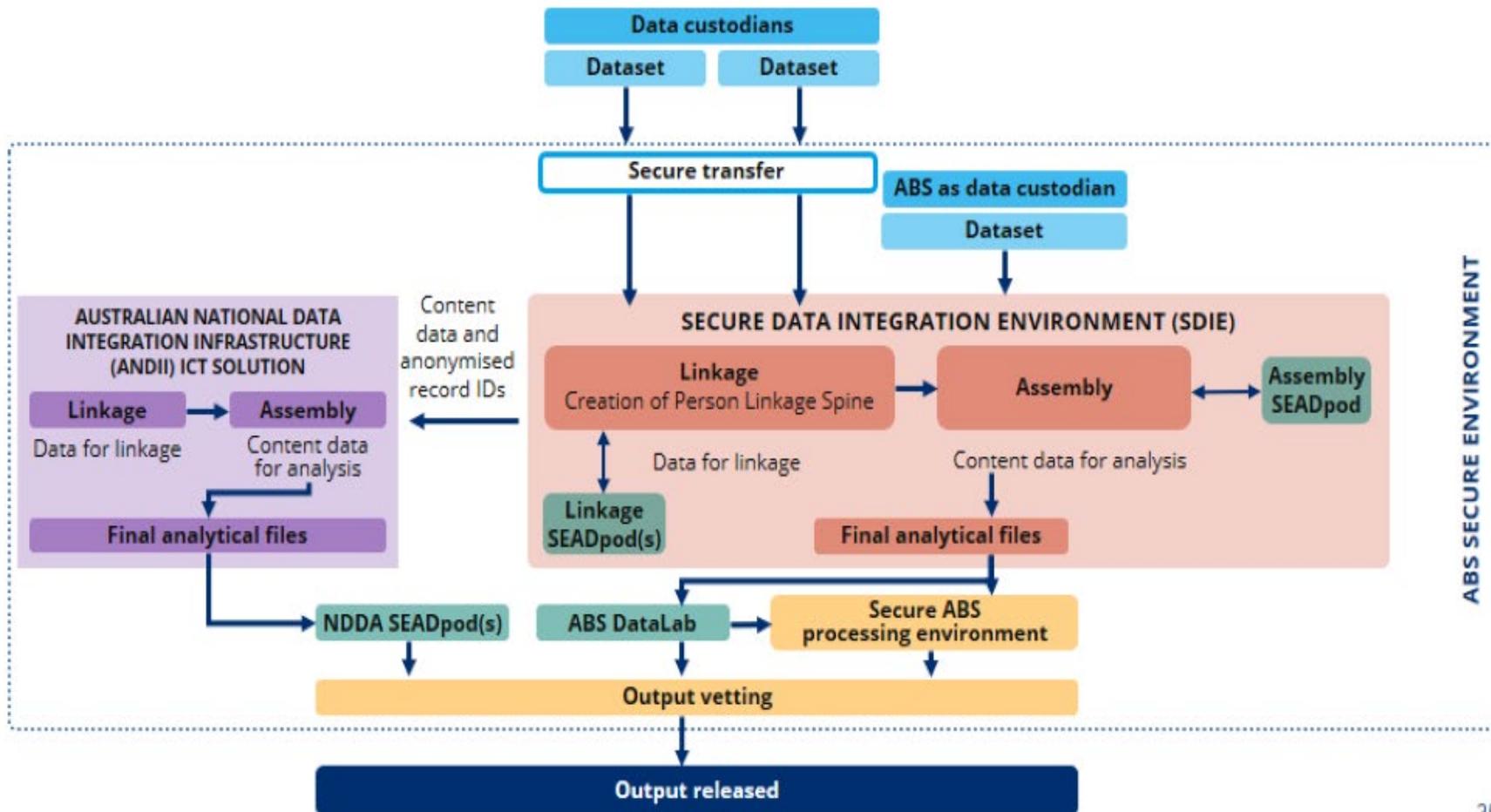
- ▶ **Proposal:** Remove the ABS requirement for functional separation between librarian and linker roles by combining these roles into a single functional role.
- ▶ Currently, ABS officers are assigned to specific functional roles for data integration:
 - Librarian: Prepares and anonymises identifying the data for linkage
 - Linker: Finds links between datasets using anonymised linkage data
 - Assembler: Creates files for analysis using linkage results
- ▶ Only Librarians can see linking data (i.e personal identifiers) in original form.
- ▶ Combining the librarian and linker role means that a slightly larger group of trusted individuals will have access to linking data in its original form.
- ▶ **Benefit:**
 - Separation principle is maintained as linkage and analytical data kept separate at all times
 - Better linkage outcomes and quality improvements
 - More efficient process to enable information to pass from Linker/Librarian's to assemblers

Changing implementation of functional separation

- Case study 4



Re-use of PLIDA data for the NDDA and ANDII, and migrating the SDIE into ANDII – Case study 5

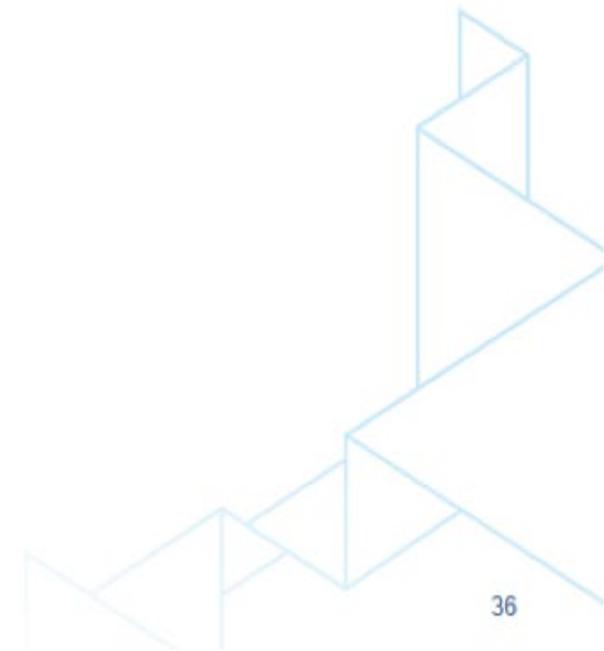


New or changed data handling practices

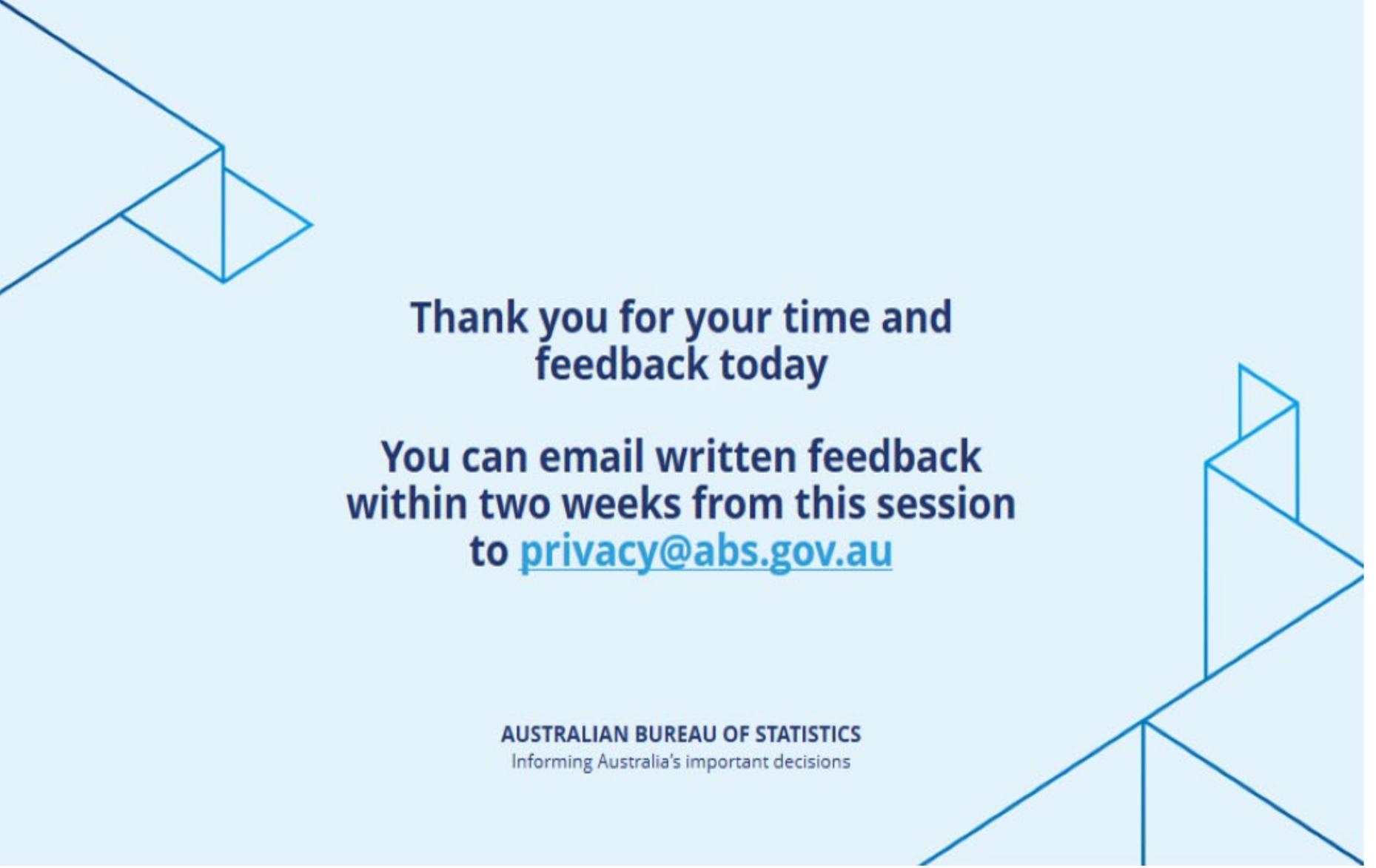


Questions for discussion:

- 7 Do you have any concerns about the proposed new or changed data handling practices?



36



**Thank you for your time and
feedback today**

**You can email written feedback
within two weeks from this session
to privacy@abs.gov.au**

AUSTRALIAN BUREAU OF STATISTICS
Informing Australia's important decisions



Principles for including new data in PLIDA



Public Benefit	The new data or data type must provide a public benefit
Utility	The new data or data type must be of sufficient quality and usefulness to be included in the asset
Sensitivities or Restrictions	New data and data types that have sensitivities are treated with additional protections or restrictions on use where appropriate
Safety	New data, data types, and handling processes are managed safely and assessed for risks

38