

Final Report

Review of Australia's Longitudinal Data System

2016



Table of contents

Acronyms.....	2
1 Executive summary.....	3
2 Context for the Review.....	6
PART I – THE VALUE OF LONGITUDINAL DATA.....	8
3 Definition of longitudinal data.....	9
4 The value of longitudinal data.....	11
5 Design and administration challenges associated with longitudinal studies	14
PART II – AUSTRALIA’S CURRENT LONGITUDINAL DATA SYSTEM	15
6 Main elements of the longitudinal data system	16
7 Strengths of the longitudinal data system.....	28
8 Weaknesses in the longitudinal data system	29
9 Opportunities for the longitudinal data system.....	37
PART III – THE PATH FORWARD FOR AUSTRALIA’S LONGITUDINAL DATA SYSTEM	38
10 The vision for the longitudinal data system.....	39
11 Recommended strategy to realise the vision for the longitudinal data system	40
12 Roadmap for implementation	58
Appendix A Terms of Reference.....	60
Appendix B Steering Committee and stakeholder consultations.....	62
Appendix C Examples of the value of longitudinal data	63
Appendix D Catalogued longitudinal datasets	66
Appendix E Key gaps in Australia’s Core National Longitudinal Data Assets.....	67
Appendix F International case studies	69
Appendix G Glossary of terms.....	70

Acronyms

Acronym	Name	Acronym	Name
ABS	Australian Bureau of Statistics	GFC	Global Financial Crisis
ACARA	Australian Curriculum, Assessment and Reporting Authority	HEIMS	Higher Education Information Management System
ACLD	Australian Census Longitudinal 5% Dataset (2006 and 2011)	HILDA	Household, Income and Labour Dynamics in Australia
AIFS	Australian Institute of Family Studies	IDI	New Zealand's Integrated Data Infrastructure
AIHW	Australian Institute of Health and Welfare	JASON	The Longitudinal Dataset for the Investment Approach
ALSMH	The Longitudinal Study on Male Health	LSAC	Longitudinal Study of Australian Children
ALSWH	The Longitudinal Study on Women's Health	LSAY	Longitudinal Surveys of Australian Youth
APS	Australian Public Service	LSIC	Longitudinal Study of Indigenous Children
APSC	Australian Public Service Commission	MADIP	Multi-Agency Data Integration Project
ATO	Australian Taxation Office	MURFs	Main Unit Record Files
BETA	The Australian Government Behavioural Economics Team in the Department of the Prime Minister and Cabinet	NAPLAN	National Assessment Program - Literacy and Numeracy
BLADE	Business Longitudinal Analytical Data Environment	NCLD	National Centre for Longitudinal Data
BNLA	Building a New Life in Australia – The Longitudinal Study of Humanitarian Migrants	NHMRC	National Health and Medical Research Council
CATI	Computer-Assisted Telephone Interviewing	NISA	National Innovation and Science Agenda
CCRC	Cannabis Cohorts Research Consortium	NPP	New Policy Proposal
COAG	Council of Australian Governments	NZDFP	New Zealand Data Future Partnership
DET	Department of Education and Training	PM&C	Department of the Prime Minister and Cabinet
DHS	Department of Human Services	RADL	ABS Remote Access Data Laboratory
DSS	Department of Social Services	SURE	SAX Institute Secure Unified Research Environment
		The Review	Review of Australia's Longitudinal Data Architecture
		VET	Vocational Education and Training

1. Executive summary

Context for the Review

The Australian Government is committed to harnessing the power of data as a strategic national resource. Optimising Australia's longitudinal data system is essential to this agenda.

Longitudinal studies follow the same participants over time. They enable researchers to look at the differences between individuals, households or entities, and how they change over time. This data can provide powerful insights about the life-course and can greatly enhance the public policy making process.

Longitudinal data also enables policy makers to understand which interventions work by studying the impact of certain actions and comparing before and after effects. Researchers can control for factors to isolate causal relationships from confounding variables such as social background and broader trends in society. In this way, longitudinal data offers much greater insights into Australian life and society than cross-sectional data which only provides a point-in-time 'snapshot'.

The Review of Australia's Longitudinal Data System (the Review) was tasked with examining Australia's future longitudinal data needs and the system that should support this data. The Review has sought to identify the strengths, weaknesses and opportunities in Australia's current approach to longitudinal data, and identify priorities to strengthen the system and realise the full potential of longitudinal data.

Strengths, weaknesses and opportunities

Australia's longitudinal data assets have developed in an ad-hoc and disconnected way. The system is largely diffuse with relatively little whole-of-system governance and coordination. Despite this, Australia's longitudinal data assets have delivered significant public value.

They provide a wide range of insights into the Australian life-course and various public policy issues. There is increasing innovation in the use of longitudinal data, including through the work of various initiatives to link administrative datasets. This system has also successfully engendered public trust, including through strong privacy protections.

Despite these strengths, a string of system weaknesses are hampering the full potential of longitudinal data to enhance our understanding of Australian life and enrich the policy making process. Six main weaknesses were identified in the Review's consultations:

1. The potential value of Australia's longitudinal data is not being realised.
2. Australia's planning for and investment in longitudinal data is uncoordinated and short-term.
3. Longitudinal data is often collected, documented, stored and accessed inconsistently.
4. It is often difficult for researchers and policymakers to access and analyse longitudinal data.
5. There is insufficient linking of longitudinal survey and administrative data.
6. There is insufficient capability to analyse longitudinal data and translate its insights.

A consistent theme throughout the Review's consultations was that this is the right time for significant improvements to be made to Australia's longitudinal data assets and system to overcome these weaknesses. The public data agenda has much momentum and there is a growing recognition of the value of public data as a national asset. In addition, advancements in technology are creating new possibilities to increase the capacity and capability of government to create, link and analyse data in a way that maintains privacy and confidentiality. Australia is also able to look to its international counterparts to learn from their experiences in collecting and using longitudinal data.

The vision for Australia's longitudinal data system

The Review envisages a future Australian longitudinal data system in which longitudinal data drives improved lifetime wellbeing for Australians. Australia's longitudinal data system should be world-class in providing a rich picture of our population and businesses. It should deliver timely, practical and relevant information to improve policies and services to citizens – reaching across enduring policy themes such as health, education, employment, society and community.

This vision will be achieved by fostering a system that works effectively across the longitudinal data value chain, from planning the development of longitudinal datasets to the policy application of the insights generated.

Recommended actions to achieve the vision

The Review's recommended priorities to achieve the vision for Australia's longitudinal data system are centred on the themes of preserve, strengthen and invest.

1. Preserve existing longitudinal data assets with sufficient funding to realise their full value

There is significant value in Australia's existing longitudinal data assets. Longitudinal data realises its greatest value over time and therefore requires a stable, long-term funding base. Australia should learn from the experience of its international counterparts on the importance of preserving investments in longitudinal data.

The immediate priority should be to preserve and unlock the value of Australia's existing longitudinal data assets, starting with the initial 15 Core National Longitudinal Data Assets identified by the Review. These are made up of eight ongoing surveys and seven linked or 'curated' administrative datasets that the Review has nominated as central to Australia's national base of longitudinal data. Other longitudinal data assets will likely be added to this base over time.

This requires a coordinated approach to the development of a funding model which can sustain and optimise these longitudinal data assets. This may include exploring co-funding mechanisms and partnerships, as well as options to improve the efficiency and operation of the data assets.

2. Strengthen the longitudinal data system

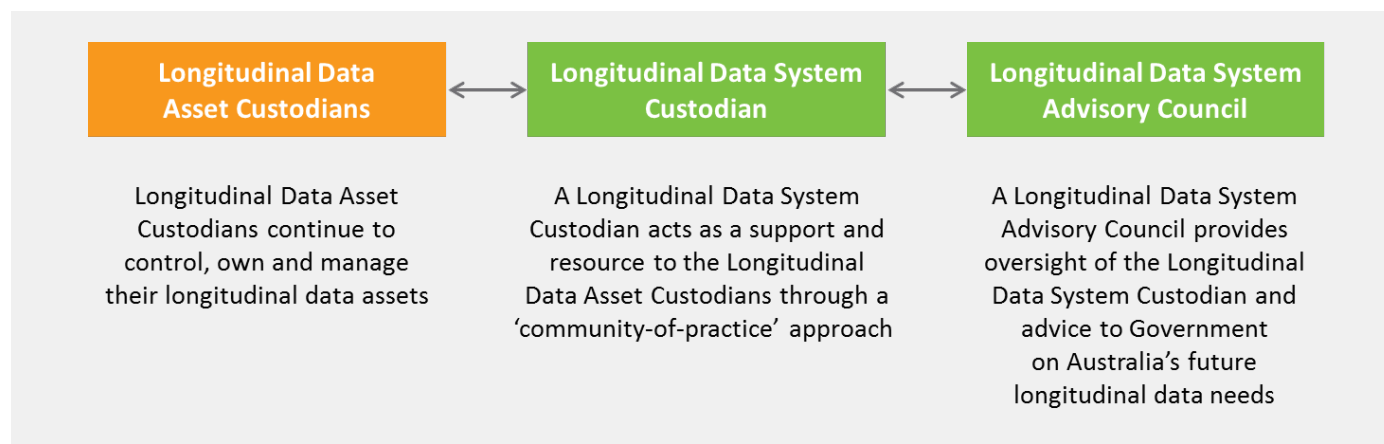
One of the Review's key findings is that Australia's current longitudinal data system is fragmented and uncoordinated. There is a need to strengthen this system through enhanced governance and coordination. The Review recommends that this be achieved through a federated governance model, whereby longitudinal data asset owners continue to manage their assets but are supported by a Longitudinal Data System Custodian and an Advisory Council. These bodies would help to facilitate and promote a consistent approach to longitudinal data planning, investments and standards on an opt-in basis. The proposed governance arrangements are shown in the diagram below.

There is also a need to foster an environment that optimises access to, and linkage of, Australia's longitudinal data assets. This requires addressing various barriers related to privacy, best practice standards for metadata, data access protocols, data linkage arrangements and data access platforms.

3. Invest in the harmonisation, coverage and impact of longitudinal data

Once the longitudinal data system has been strengthened, the priority should be to expand the coverage, depth and impact of Australia's longitudinal data assets. This may include considering options to fill gaps in the life-course coverage of Australia's longitudinal data, such as the early years, Indigenous perspectives, interpersonal violence, and retirement and ageing.

There should also be a focus on investing in the promotion and application of longitudinal data through projects that demonstrate its value and potential use in the decision-making process. These projects should demonstrate the power of longitudinal data across the value chain – from linking data, to robust analysis, to presenting policy relevant results in compelling ways.

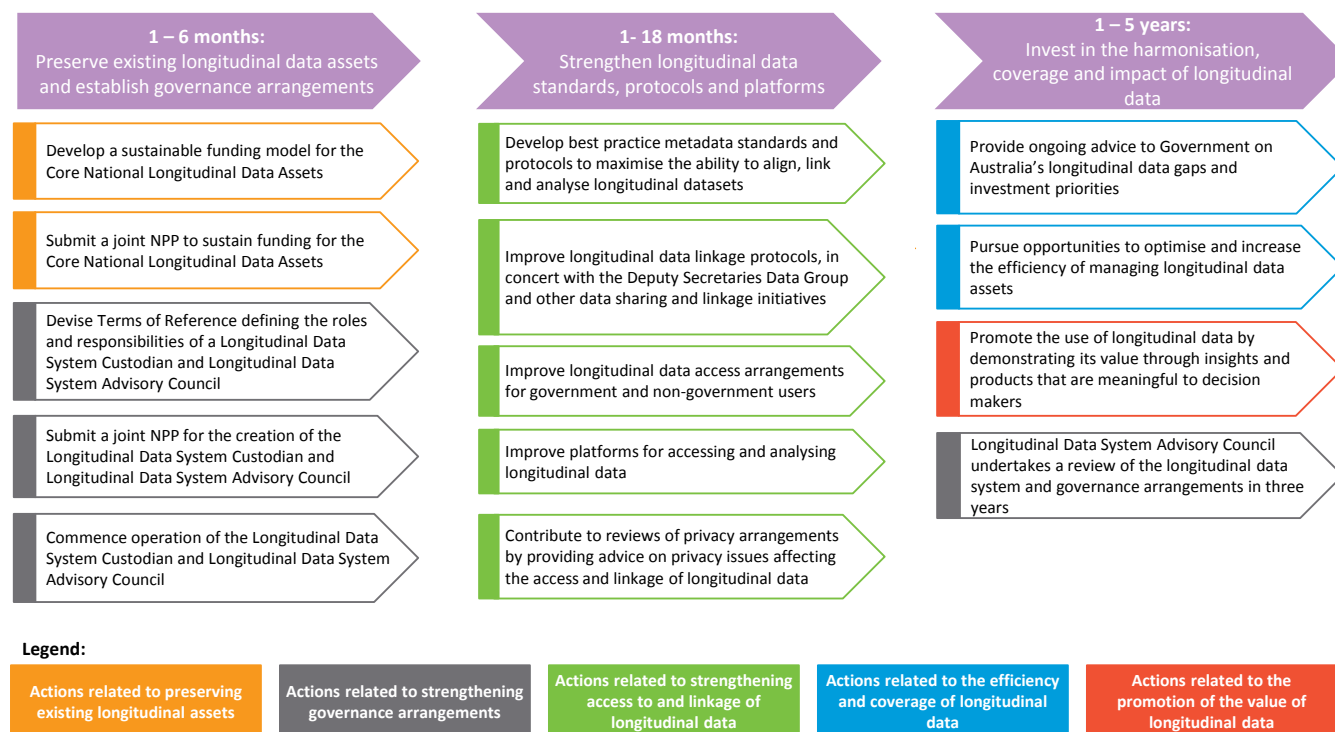


Next steps

The Review suggests that an implementation working group be established to advance the recommended priorities in this report, with a particular focus on preserving funding for the Core National Longitudinal Data Assets and establishing the proposed governance arrangements for the longitudinal data system. This may require the development of a joint New Policy Proposal (NPP) to the Australian Government.

The Review considers that the National Centre for Longitudinal Data (NCLD) is well placed to play the role of the initial Longitudinal Data System Custodian during this establishment phase. The implementation working group, organised by the NCLD, should devise a Terms of Reference to agree on these governance and funding arrangements.

IMPLEMENTATION ROADMAP



2. Context for the Review

The Government is committed to optimising the use of data

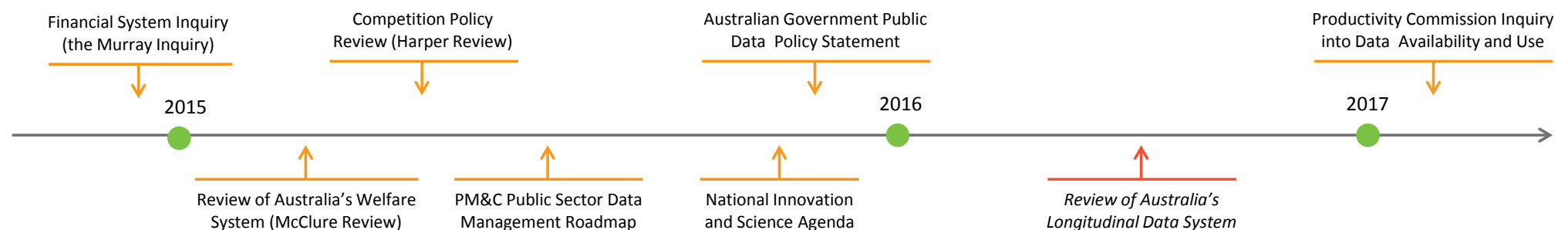
Harnessing the value of data is a national priority for the Government. The volume and scope of data is growing exponentially and so are the opportunities for its application by government, business, academia and citizens.

The *Public Data Policy Statement* recognises that Australia's capacity to remain competitive in an increasingly digital economy is contingent on its ability to make full use of the value of public data. The data held by government is a strategic national resource that holds considerable value for growing the economy, improving service delivery and transforming policy outcomes for the Nation.

The *Public Sector Data Management Report* similarly acknowledges the critical role of data in fostering innovation. The right data at the right time can be used to make better and faster decisions, encourage new efficiencies and increase productivity.

This Review sits within a broader landscape of reviews and initiatives focussed on helping Australia to realise the value of data as a strategic national resource. This includes the following:

- Financial System Inquiry (2014): recommended the Government task the Productivity Commission to review the benefits and costs of increasing the availability and use of data.
- McClure Review of the Welfare System (2015): recommended the integration of real time data for income support reforms, as well drawing on longitudinal data.
- Harper Review of Competition Policy (2015): recommended that the Government consider ways to improve individuals' ability to access their own data to inform consumer choices.
- Public Sector Data Management Report (2015): a cross-agency review led by the Department of the Prime Minister and Cabinet, and charts a course for optimising public sector data to achieve efficiencies for government, enable better service delivery and stimulate economic activity.
- National Innovation and Science Agenda (NISA) (2015): committed the Government to releasing more non-sensitive public data for private sector innovation, and making use of data to improve service delivery and inform policy, including by funding Data61 – Australia's largest data innovation group.
- Public Data Policy Statement (2015): committed the Government to optimise the use of public data, to release non-sensitive data as open by default, and to collaborate with the private and research sectors to extend the value of public data.
- Productivity Commission Inquiry on Data Availability and Use (due for release in 2017): a broad ranging investigation into the benefits and costs of options for improving the availability and use of data.



Longitudinal data is a key part of the public data agenda

Optimising Australia's collection of longitudinal data is essential to realising the full value of data as a strategic national resource. Longitudinal data has the ability to provide powerful insights about Australian life and society that cannot be obtained from other data. While cross-sectional data can provide a useful 'snapshot' of what is happening in Australia, only longitudinal data can provide insights into individual changes, transitions and trajectories across the life-course, as well as insights on causation.

The Australia Government has an extensive suite of longitudinal data, and makes a significant investment in longitudinal surveys. For example, the Department of Social Services (DSS) funds four longitudinal surveys at a cost in excess of \$20 million per year. The Government is also dedicating resources to developing administrative data collections that will provide a longitudinal perspective. There are also increasing efforts to link survey and administrative data to further enhance these insights.

These data assets are especially valuable as an input to the 'Investment Approach' that the Australian Government is applying to the social security system, consistent with the recommendations of the McClure Review of Australia's Welfare System.

This work will be used to guide policy development and interventions to increase people's capacity to live independently of welfare, and to address the risk of intergenerational welfare dependency. The 'Investment Approach' will continue to be heavily dependent upon longitudinal data to generate accurate and robust estimates of the liability, as well as the potential reduction in the liability from policy interventions and improvements in outcomes.

To advance this agenda, the Australian Government commissioned this Review of Australia's Longitudinal Data System to inform Australia's future longitudinal data needs as part of the "Investment Approach to Welfare" measure in the 2015-16 Budget.

This approach also has significant implications for the health system. Longitudinal data can be used to increase the evidence base on physical and mental health, and better target health initiatives related to prevention, education and treatment.

Longitudinal data on firms is also important for assessing the impact of government policies, including those aimed at stimulating innovation. Datasets that are linked to employment information can provide guidance on a range of policy issues, such as which investments in education and training would best support economic growth.

The Review was asked to explore options for Australia's future longitudinal data system

The strategic objectives of the Review were to:

- develop a strategic plan for the Australian Government's longitudinal data architecture to support evidence-based policy development
- promote a coordinated approach to longitudinal data developments, informed by the needs of government, academia and the private sector
- consider arrangements for supporting longitudinal data integrity, development, analysis and dissemination in Australia
- consider the current and potential use of longitudinal administrative data, the complementary role of longitudinal surveys, and how to best balance investments in these data assets
- propose a framework prioritising investment by the Commonwealth in new and existing longitudinal survey and administrative data assets.

See Appendix A for the Review's Terms of Reference.

The Review has been overseen by a whole-of-government Steering Committee that provided strategic advice and guidance. The Review also engaged with a wide range of stakeholders from the Commonwealth and State Governments, research institutes, universities, think tanks, the private sector and international counterparts (see Appendix B for a full list of Steering Committee members and consultations).

The operation of the Review has been managed by the NCLD within DSS. The NCLD was established to bring together the longitudinal surveys managed by DSS. Its mission is to promote a longitudinal evidence base that informs policies and practices to improve the lifetime wellbeing of people and families in Australia.

PART I THE VALUE OF LONGITUDINAL DATA

3. Defining longitudinal data

Longitudinal data follows the same participants, often over the course of many years. It allows researchers to look at the differences between individuals, households or entities, and how they change over time. This can provide powerful insights about trajectories and variations in outcomes over the life-course. Examples of longitudinal datasets are shown to the right.

At the broadest level, there are two types of longitudinal data: designed data and organic data. Designed data is usually generated from surveys of the same individuals at different points in time, while organic data is the by-product of registrations, transactions and record keeping by government (usually called 'administrative data') or the private sector.

Longitudinal data can also be linked to derive more powerful and robust insights. This can occur in three main ways:

- **Linking administrative data:** government datasets can be linked to create a longitudinal picture. For example, DSS has created 'JASON', a dataset containing the characteristics and payment entitlements history of income support recipients.
- **Linking survey and administrative data:** longitudinal survey participants can consent to have their responses linked with administrative datasets. For example, the Longitudinal Study of Australian Children (LSAC) links to several administrative databases, including Medicare data and NAPLAN.
- **Linking surveys:** surveys can be linked to create or enhance the longitudinal picture. For example, the ACLD links a 5% random sample from the 2006 and 2011 Census (see the case study box).

HOUSEHOLD, INCOME AND LABOUR DYNAMICS IN AUSTRALIA SURVEY

In 2001 the Household, Income and Labour Dynamics in Australia (HILDA) Survey started interviewing a sample of individuals within Australian households about their economic and subjective wellbeing, labour market dynamics and family dynamics. Each year, researchers have followed up with these households and tracked the changes (and stability) in their lives.

HILDA has produced a wide range of insights about Australian life, including household and family relationships, employment, child care, education, income, expenditure, health and wellbeing, attitudes and values on a variety of subjects, and various life events and experiences.

BUSINESS LONGITUDINAL ANALYTICAL DATA ENVIRONMENT

The Business Longitudinal Analytical Data Environment (BLADE) is a longitudinal administrative data asset managed by the Department of Industry and Science and the Australian Bureau of Statistics (ABS). The BLADE integrates tax activity, business characteristics and economic activity data from the ABS and the Australian Taxation Office (ATO) to provide firm-level data on over two million active businesses in Australia. The BLADE is a powerful tool for informing policy and evaluating programs related to employment and economic growth. For example, research using data from the BLADE recently found that start-ups are the highest contributor to net job creation. This in turn has informed the Australian Government's National Innovation and Science Agenda (NISA).

THE LONGITUDINAL SURVEYS OF AUSTRALIAN YOUTH

The Longitudinal Surveys of Australian Youth (LSAY) follows young people as they move from school into further study, work and other destinations. LSAY uses large, nationally representative samples of young people to collect information about education and training, work, and social development. LSAY provides a rich source of information to help better understand young people and their transitions from school to post-school destinations, as well as exploring social outcomes, such as wellbeing.

AUSTRALIAN CENSUS LONGITUDINAL DATASET

Australian Census Longitudinal Dataset (ACLD) uses data from the Census of Population and Housing to build a longitudinal picture of Australian society. In this first release, a 5% random sample from the 2006 Census was brought together with records from the 2011 Census using data linkage techniques without names and addresses. In taking a longitudinal view of Australians, the ACLD may uncover new insights into the dynamics and transitions that drive social and economic change over time, including how these vary for diverse population groups and geographies.

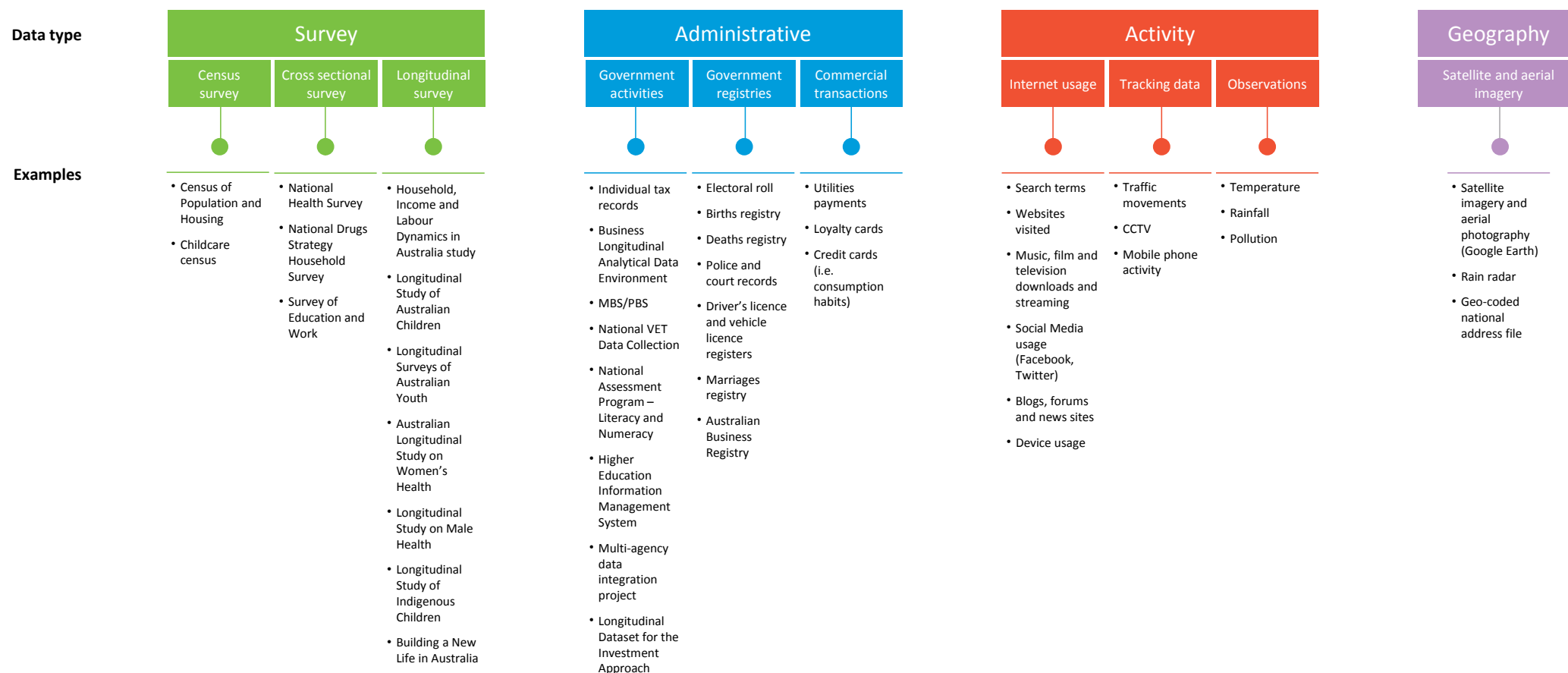
Longitudinal data is one part of the expanding data landscape

The amount of data being generated by individuals, government and companies is increasing exponentially. More data has been created in the past two years than in the entire previous history of the human race. Best estimates suggest that at least 2.5 quintillion (one quintillion equals a billion billion) bytes of data are produced every day.

While longitudinal data has traditionally been confined to surveys designed to track a small sample of participants by asking them a small number of questions at regular

intervals, technology is enabling us to make use of the much wider and more regular coverage offered by the vast amount of data collected by governments and firms. The construction of a longitudinal picture can no longer be thought of within the limits of longitudinal surveys, but instead must be considered as a feature that spans the entire data landscape.

An overview of this expanding data landscape is shown in the diagram below.



4. The value of longitudinal data

Data helps to inform an understanding of Australian life

Data is crucial to informing our understanding on Australian life and society. This understanding can have applications for a wide range of actors, including governments, researchers, the private sector and citizens. It can be particularly important for the public policy process.

Evidence-based policy is in strong demand to help governments achieve their desired outcomes in the most effective and efficient way possible. This has been emphasised as part of the Australian Government's Public Data Policy Statement and Public Sector Data Management Report.

The first step is for policymakers to be able to understand what is happening in society and why it is happening. Once the desired outcomes have been set, government needs to know which policies and programmes will be likely to achieve these outcomes. Policymakers must then be able to evaluate interventions and determine whether they achieved the desired outcomes in an effective and efficient way.

Each stage of this process is heavily dependent on data. Without data, policymakers are left to rely on intuition, anecdotes and conventional wisdoms. Data can help overcome these tendencies by equipping policymakers with a more precise view of a particular problem and pointing them towards potential solutions.

Cross-sectional data provides useful but limited insights

A large amount of public data is cross-sectional: studies that observe a sample of the population (such as individuals, households or businesses) at a single point in time. This offers a 'snapshot' of a particular population.

Cross-sectional studies can also be repeated by asking the same questions at different time-points to a new sample of people on each occasion.

Cross-sectional data is very good at providing two types of insights: level and trend insights. Cross-sectional data can tell us how many people are experiencing a particular event or circumstance at a certain point in time (level). It can also tell us about aggregate changes in both the number and characteristics of people in a particular circumstance (trend).

These sorts of insights are useful because they provide an accurate picture of what is happening in society at an aggregate level. However, cross-sectional studies have two main limitations: they offer limited insight into individual changes and transitions, and are not very good at explaining why something has occurred.

For example, a snapshot of the poverty rate from one year to the next might show little change. However, this does not tell us whether the same people who were in poverty in the first year were the same people in poverty in the second year, or whether new people entered into poverty. While cross-sectional studies can ask respondents to recall their past circumstances, the reliability of these answers is usually limited. In other words, cross-sectional studies can only provide measures of *net* rather than *gross* change.

Longitudinal data provides insights on individual change and causation

Like cross-sectional studies, each wave of a longitudinal study provides a snapshot of the population of interest. Accordingly, it is possible to use each wave of a longitudinal study to derive level and trend insights.

The real value of longitudinal data comes from its ability to produce two additional types of insights: individual change, transition and trajectory insights, and causation insights. Longitudinal data can tell us where people come from before they enter a particular circumstance, how long people with certain characteristics tend to spend in a particular circumstance, where people go after they leave a particular circumstance, and whether and how often people return to a particular circumstance.

Longitudinal data on firms enable researchers to answer questions such as whether investments in research and development or training improves the performance of firms over time, and whether this is true for all types of firms or only some. It can provide insights such as why some firms adopt new technologies while others do not, and whether this affects employment and wages.

Longitudinal data enhances our understanding of which interventions work by studying the impact of a certain action and comparing before and after effects on individuals, households and firms. Researchers can control for factors to isolate causal relationships from confounding variables such as social background and broader trends in the economy and society.

To return to the example of poverty, a longitudinal study can tell us how long people with certain characteristics tend to spend in poverty. This lets us know whether the experience of poverty tends to be brief and isolated, or long-term and recurring. A longitudinal study can also examine the factors and characteristics that tend to precede a person moving into poverty. Together this starts to build a picture of what causes poverty and which interventions help to reduce it.

See Appendix C for more detailed examples of the sorts of insights that can be generated from longitudinal data.

Insights provided by both cross-sectional and longitudinal studies

LEVEL INSIGHTS

Both cross-sectional and longitudinal studies can tell us how many people are experiencing a particular event or circumstance at a certain point in time. They can also tell us about their characteristics.

For example, every five years the Census provides a snapshot of the number and key characteristics of people who are in Australia and their housing. This information provides a reliable basis for estimating the population, which can be used for purposes such as defining electoral boundaries and distributing government funds.

While level insights can be provided by both cross-sectional and longitudinal studies, the cost and administrative burden of tracking respondents over time often means that longitudinal studies have smaller sample sizes than cross-sectional studies. Consequently, cross-sectional studies sometimes offer more accurate estimates of what is occurring in the broader population of interest than longitudinal studies.

TREND INSIGHTS

By definition longitudinal studies provide insights into trends because they track the same participants over time. Cross-sectional studies also offer trend insights when they are repeated with a new sample. Such studies provide insights about aggregate changes in both the number and characteristics of people and firms in a particular circumstance.

For example, the National Drug Strategy Household Survey collects information on alcohol, tobacco and illicit drug use in Australia and has been conducted every 2 to 3 years since 1985. It surveys different people each time and tries to ensure that each new sample is representative of the broader Australian population.

To take smoking as an example, the 2013 survey found that 12.8% of people aged 14 or older smoked on a daily basis, and people living in remote areas were about twice as likely to smoke as those living in cities (level insights). By comparing these results to previous years, the survey showed that from 2010 to 2013 there was a significant decline in the proportion of the population who smoke on a daily basis, but there was no significant reduction in the proportion of people from regional areas who smoke over the same period (trend insights).

Insights only provided by longitudinal studies

INDIVIDUAL CHANGE, TRANSITION AND TRAJECTORY INSIGHTS

Longitudinal data offers a deeper level of analysis by measuring change over time at an individual level. For example, the National Drug Strategy Household Survey shows that the proportion of the population that smokes on a daily basis has declined. But it does not tell us how long people with certain characteristics are smokers, when they quit and whether they resume smoking.

Longitudinal data can also shed light on the transitions that people make in and out of particular circumstances. For example, longitudinal studies can examine the factors and characteristics that tend to precede a person starting and stopping smoking. They can also show the different lifestyle transitions that people make when they stop smoking.

CAUSATION INSIGHTS

Repeated cross-sectional studies are not well suited to explaining why something has occurred. Cross-sectional studies cannot control for other factors, nor can they distinguish between correlation and causation.

The National Drug Strategy Household Survey does not give us many insights on why smoking rates have declined, be it increases in taxes on cigarettes, the introduction of plain packaging, public health campaigns or some other reason. For instance, the study points to a correlation between living in a city and being a non-smoker, but cannot tell us whether living in a city causes people to stop smoking.

Longitudinal data would help policymakers to better understand these effects and evaluate which policy interventions work. This could help better target government interventions to improve both the effectiveness and efficiency of government programmes.

INSIGHT COMPARISON

The figure to the right provides a comparison of the value of cross-sectional vs longitudinal data.

Cross-sectional data		vs	Longitudinal data	
LEVEL INSIGHTS				
The number of people in a particular circumstance	Larger and broader samples can provide more precise estimates of prevalence	✓	✓	The precision of estimates can be limited by the need for smaller and narrower samples in surveys
The characteristics of the group of people in a particular circumstance		✓	✓	
AGGREGATE TREND INSIGHTS				
The number of people in a particular circumstance over time	Requires repeated cross-sectional data	✓	✓	The sample may become less representative of the target population over time
Changes in the characteristics of the group of people in a particular circumstance over time		✓	✓	
INDIVIDUAL CHANGE, TRANSITION AND TRAJECTORY INSIGHTS				
Where people come from before they enter a particular circumstance	Not unless relying on respondent recall or linkages to historical administrative data	✗	✓	Survey responses are recorded during or close to the event, and can be supplemented and validated by linked administrative data
The length of time people spend in a particular circumstance		✗	✓	
Where people go to after they leave a particular circumstance		✗	✓	
Whether and how often people return to a particular circumstance		✗	✓	
CAUSATION INSIGHTS				
The factors that cause a particular circumstance	Very difficult to control for other factors and distinguish between correlation and causation	✗	✓	Tracking variables provides a greatly enhanced capacity to separate correlation from causation

5. Design and administration challenges associated with longitudinal studies

While longitudinal studies provide much deeper insights than cross-sectional data, they are also usually more difficult and costly to design and administer. This is especially true when longitudinal studies depend on surveys of individuals, households or firms. These issues can be broadly classified into design challenges and administration challenges.

DESIGN CHALLENGES

Most of the design challenges associated with longitudinal surveys are derived from the difficulty of tracking people, households or firms over a long period of time. Consequently, longitudinal surveys tend to use a smaller sample than cross-sectional studies and often study a narrower population of interest. There are also usually heightened privacy and ethical concerns associated with tracking individuals over a long period of time, particularly if the study links with administrative data.

The set sample in a longitudinal survey also makes it difficult to continue to represent the population of interest over time. The sample is usually set in the first wave and may not keep up with broader changes in the population of interest, such as migration patterns.

Similarly, longitudinal studies carry the risk that participants will, for a variety of reasons, drop out of future waves of the survey. This attrition can skew the data, either because the sample no longer accurately represents the population of interest or because some participants are more likely to drop out than others. The application of population weights can only partly alleviate this problem. In such cases it may be necessary to refresh or top-up the sample to maintain its representativeness.

There is also the risk of 'panel conditioning' whereby participants' responses are affected because they have been part of the study for multiple waves. While this can also occur for cross-sectional studies, the repeated and long-term nature of longitudinal studies increases the risk.

ADMINISTRATION CHALLENGES

The heaviest administrative burden of longitudinal studies comes from the volume and complexity of data that is collected. This presents data management, data storage, staffing and funding challenges. Running a longitudinal survey of the breadth and length of HILDA or the British birth cohort studies is an enormous endeavour that depends on a long term commitment from a wide range of stakeholders.

The total cost of a longitudinal study across multiple waves is usually much larger than a cross-sectional study, due to the need to invest in participant retention and tracking initiatives. However, each wave of a longitudinal study may not necessarily cost more than a cross-sectional study. In any case, the need for long-term funding presents a recurring challenge for most longitudinal studies.

The long time-frame associated with most longitudinal studies also present a barrier. It can take many years for worthwhile results to accumulate from a longitudinal study, even though each wave provides a snapshot that can be analysed like a cross-sectional study.

Significant technical capabilities are also required to get the most value out of longitudinal data. Analysing longitudinal data requires very sophisticated statistical techniques to distinguish between correlation and causation. There is also much complexity associated with the weighting and estimation of longitudinal survey files. Retaining supply of this capability is becoming more challenging as the global demand for data experts grows.



PART II AUSTRALIA'S CURRENT LONGITUDINAL DATA SYSTEM

6. Main elements of the longitudinal data system

Australia's longitudinal data system can be examined through different lenses. It is important to note that the current system is not very well defined and operates with a variety of different governance arrangements, projects and stakeholders. These perspectives collectively provide an overview of the current approach to longitudinal data. These system elements encompass:

- **Governance and privacy arrangements** – the Australian longitudinal data system operates within a diverse and often complex web of arrangements relating to coordination, decision-making and the protection of privacy.
- **Longitudinal data value chain** – each longitudinal dataset (or datasets where there are linkages) is the product of a 'value chain' that spans five phases and is enabled by technology: data planning, data collection, data custody and access, data analysis and research, and end-use.
- **Main actors** – there is a diverse range of government and non-government actors who perform roles across the longitudinal data value chain, from funding longitudinal studies through to the application of the resulting analysis to the public policy making process.
- **Core National Longitudinal Data Assets** – the Review has identified 15 initial 'Core National Longitudinal Data Assets'. These comprise eight ongoing surveys and seven linked or 'curated' administrative datasets.

Each of these system elements is explored in this section.

6.1 Governance and privacy arrangements

The longitudinal system operates through a variety of arrangements relating to governance and privacy. The main elements are summarised below.

Governance arrangements are diffuse but there are some initiatives underway to improve coordination

Australia's longitudinal data system is comprised of many actors working across multiple sectors and jurisdictions. The ABS is enabled by legislation to act as the central statistical authority for the Australian Government and provider of statistical services for the states. This legislation charges the ABS with responsibility to "ensure co ordination of the operations of official bodies in the collection, compilation and dissemination of statistics and related information" with particular regard to "the avoidance of duplication", "the attainment of compatibility and... the integration of statistics" and "the maximum possible utilization" of statistics (*Australian Bureau of Statistics Act 1975*, s 6).

In practice, no single entity currently has the resources or authority to coordinate the entire longitudinal data system. Instead, various governance mechanisms have formed to promote cooperation and data-sharing. For example, in 2010 Australian Government Portfolio Secretaries endorsed seven high level principles for the integration of Commonwealth data for statistical and research purposes as well as a supporting set of governance and institutional arrangements. They set out a common framework for statistical data integration involving Commonwealth data to take place in a safe and effective environment. A cross-portfolio 'Data Integration Oversight Board' has been established to oversee this framework and the development of a cross government environment for data integration.

Other initiatives have been established across the federation. For example, the NSW Government is currently working with the Australian Government Department of Human Services (DHS) to examine issues associated with troubled families and have developed data sharing arrangements to assist this process.

Various measures have also been taken to improve governance, coordination and innovation at the Commonwealth level as part of the broader focus on improving public sector data management. This role is primarily being performed by the Deputy Secretaries Data Group, which is charged with implementing the roadmap of recommendations from the Public Sector Data Management Report to optimise public data (longitudinal and other) to improve policy development, government service delivery and private sector innovation. An example of this sort of innovation is Data61, Australia's largest data innovation group.

DATA61 - CREATING A DATA-DRIVEN FUTURE, TODAY.

As part of NISA, the Government is investing \$75 million in Data61, Australia's largest data innovation group. Data61 is working on a number of high-value, transformative data projects. One of these involves development of world leading public data infrastructure that maximises discoverability and reuse of high-value open data. Another project involves the development of algorithms that enable the safe release of synthetic representations of a dataset covering 15 years of social security and family payments (linked at the individual level).

Multiple instruments and protocols protect privacy

There are number of existing privacy protections to ensure that the data divulged by individuals (either through organic or designed longitudinal data) generates relevant insights for researchers and policy makers whilst protecting the individual's right to privacy. These protections exist at the individual, study and regulatory levels.

At the individual level, written consent must be received from longitudinal survey participants before their administrative data can be accessed and linked. For example, an entity conducting a survey will need the permission of a survey participant before it can link their answers to their health records. In turn, the entity would need to obtain the consent of the holder of the health record for this data to be released and linked.

At the study level, the large majority of longitudinal data actors, including government entities and tertiary institutions, operate under the National Statement on Ethical Conduct in Human Research. Internal processes and terms of agreements also typically require that all longitudinal studies be approved by a Human Rights Ethics Committee to ensure ethical research standards.

At the regulatory level, there are various Commonwealth, state and territory regulations governing the collection, use and disclosure of information. While the *Privacy Act 1988* and the associated Australian Privacy Principles provide the primary framework for managing personal information, most states also have their own privacy legislation. In addition, there is often separate legislation governing management and access for specific types of data, as well as agency-specific legislation, policies and procedures for the collection, use and disclosure of administrative datasets.

6.2 The longitudinal data value chain and supporting technology

The insights provided by longitudinal data are the product of a value chain

The longitudinal data 'value chain' spans five broad phases: data planning, data collection, data custody and access, data analysis, and application of the information generated by end-users (see the figure overleaf). Every component of this value chain needs to function effectively for longitudinal data to generate value – from planning what longitudinal data will be collected through to analysing and applying its insights.

One of the most complex phases of the longitudinal data value chain is data custody and access – this is where data is extracted from either administrative or survey datasets, and prepared to be made available for data analysis. Much of the longitudinal data system is focussed on ensuring data is prepared and accessed in a way that minimises the risk of privacy and confidentiality being breached.

For example, the data processing component includes data being de-identified. This may comprise identifiers such as names, addresses and dates of birth being removed, to be replaced by data field containing age and a geographic identifier. De-identification may be sufficient for privacy protection if the data can only be accessed in a 'safe environment' by researchers who have been accredited as 'safe people'. Confidentialisation, which is required for public release, may require manipulation of the unit record data to prevent reverse engineering of the data.

It is also important to ensure that the other phases of the value chain are working effectively. The system requires careful planning to decide what longitudinal data assets to invest in, a strong capability to analyse data and translate its insights, and to apply them in decision-making.

The creation and use of longitudinal data is enabled by technology

The creation and use of longitudinal data would not be possible if not for a diverse range of enabling technologies that are applied across the longitudinal data value chain. The application of technology is a function of the underlying data characteristics. For example, sample size, method and target population group contribute significantly to the data collection method used in a survey. While some surveys can be completed through online, self-complete questionnaires, others require one-on-one interaction, such as face-to-face interviews or Computer-Assisted Telephone Interviewing (CATI).

Data custody often involves processing, storing and documenting very large datasets. Good documentation on the variables is particularly important for datasets derived from administrative data where there is not a series of survey documents that provide the questions asked. Arguably the most complex process occurs in any data linkage between administrative datasets, particularly where an existing linkage ID (e.g. Medicare number) is not present in both datasets. In this situation linkage requires sophisticated technology to undertake probabilistic matching.

Alternatively a 'backbone' matching different identifiers (such as Medicare and Tax File Numbers) can be used to provide an exact matching (at least to the extent that these identifiers are unique and stable over time). Such backbones are only ever used to assign a unique identifier to each dataset, which is then used to link the datasets so the new linked data does not contain any of the original identifiers, such as names, and addresses.

Similarly, the mode of access is often determined by the sensitivity of individual information available through the longitudinal dataset. Access regimes are highly varied, from access restricted to a project team or widely available to approved non-government researchers. Accordingly, data may be made available to researchers via a CD-ROM containing curated data or through a highly sophisticated environment such as the Sax Institute's Secure Unified Research Environment (SURE).

Additionally, analysis methods are also evolving in response to increasing analytical computer power. Today researchers can undertake statistical analysis of data containing millions of observations on their desktop computer. And web-based services such as Amazon Web Services can apply machine learning to analyse data containing billions of observations.

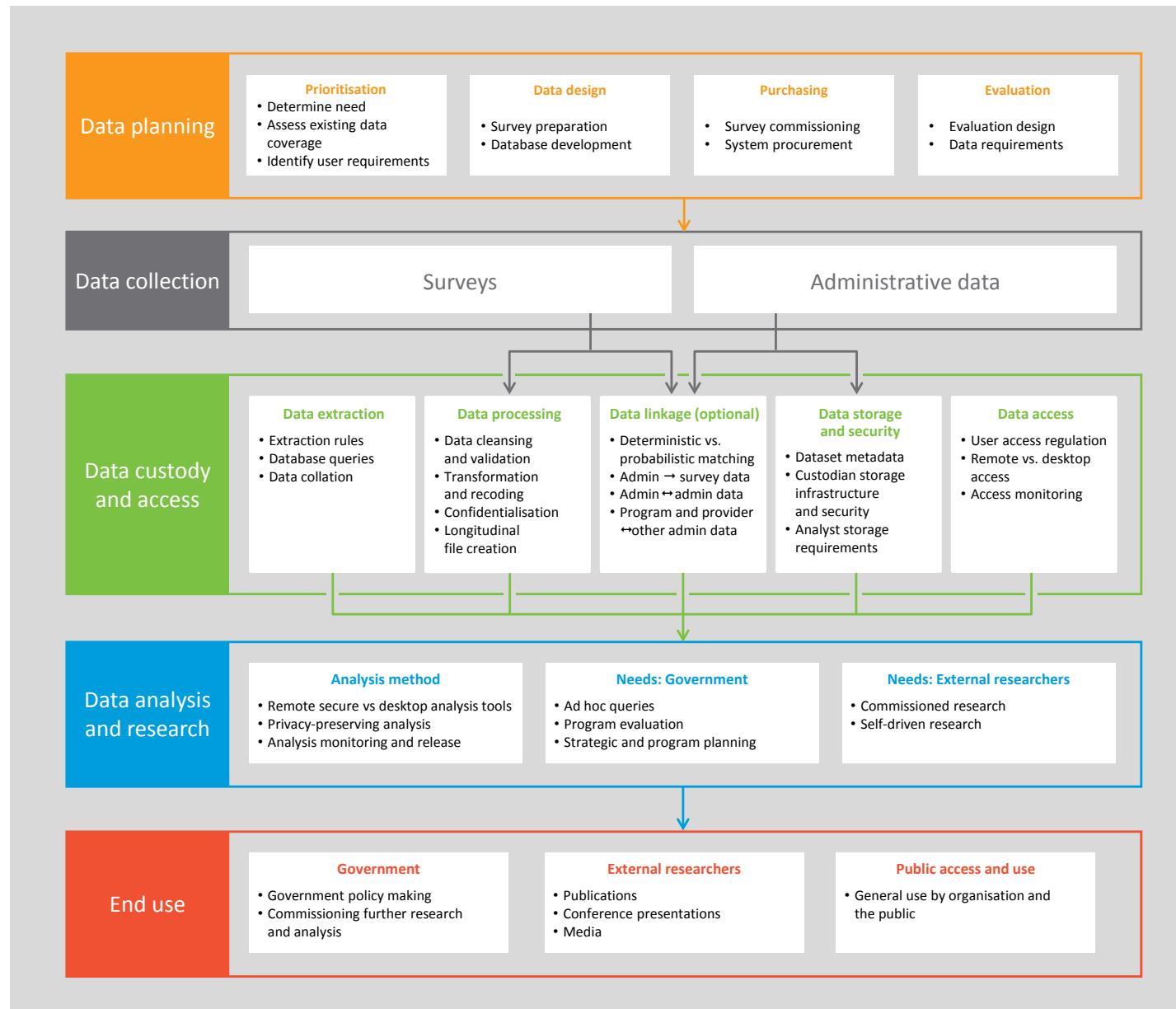
THE SAX INSTITUTE'S SURE

The SAX Institute has developed SURE – a high powered computing environment helping to bring researchers together to collaborate on large scale projects tackling major health and social issues such as population ageing, diabetes and mental health

SURE provides secure and remote virtual access to datasets over the internet. Data never leaves SURE, and cannot be copied or downloaded. Inputs and outputs are vetted, with all activities recorded and archived for future reference. Researchers can view individual-level data when using SURE, but can only extract information that cannot be used to identify individuals. In addition, strong protocols are applied to vet potential users.

THE LONGITUDINAL DATA VALUE CHAIN

The figure to the right depicts the main parts of the longitudinal data value chain.



6.3 Actors in the longitudinal data system

Australia's longitudinal data system is made up of a range of actors who perform roles across the longitudinal data value chain, from the funding of longitudinal studies through to the application of the resulting analysis to the public policy making process.

Australia's system is largely diffuse with relatively little whole-of-system governance and coordination. But there are some major actors who tend to be involved in more than one part of the value chain. For example, the ABS and DSS are involved in some capacity across all parts of the value chain.

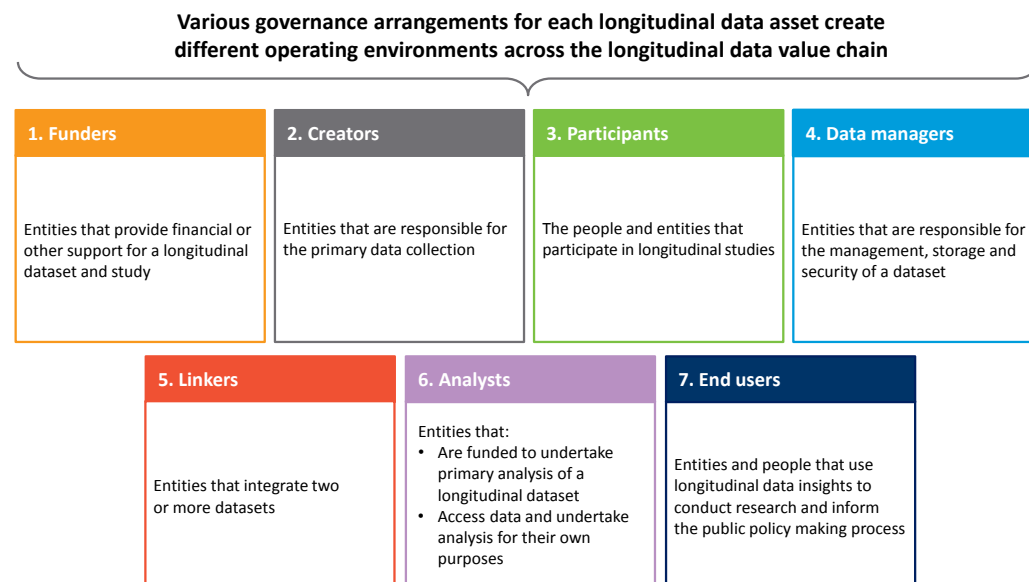
Other entities play more confined roles. For example, the National Health and Medical Research Council (NHMRC) is only a funder of longitudinal studies, while third parties such as Roy Morgan Research are contracted to collect data, but not always to analyse it.

Examples of these actors and governance arrangements for three longitudinal surveys are shown overleaf. The governance arrangements for administrative linkage projects and associated datasets are often much more complicated.

For instance, there is an important role played by data integrating authorities. The ABS, Australian Institute of Health and Welfare (AIHW) and Australian Institute of Family Studies (AIFS) are "Authorised Integrating Authorities". These entities have the authority to undertake data integration projects involving Commonwealth data for statistical and research purposes according to best practice guidelines and principles. The Deputy Secretaries Data Group is currently reviewing these accreditation arrangements. The Cross Portfolio Data Integration Oversight Board also has responsibility to help manage the systemic risk associated with conducting multiple data integration projects involving Commonwealth data.

There are also various state-based bodies that perform similar data linkage functions. This includes the Centre for Health Record Linkage (NSW/ACT), the Queensland Centre for Health Data Services, SA/NT DataLink, the Tasmanian Data Linkage Unit, Victorian Data Linkages, and the Western Australian Data Linkage Branch.

ACTORS IN THE LONGITUDINAL DATA VALUE CHAIN



EXAMPLES OF ACTORS AND GOVERNANCE ARRANGEMENTS IN LONGITUDINAL SURVEYS

Longitudinal survey	Overview	Data funders	Data creators	Participants	Data managers	Data linkers	Data analysts	End-users
Household, Income and Labour Dynamics in Australia (HILDA) study	A household-panel study that interviews households about economic and social well-being, along with labour market and family dynamics	Australian Government through DSS	The Melbourne Institute of Applied Economic and Social Research, as the manager of HILDA, contracts Roy Morgan Research to survey participants	<ul style="list-style-type: none"> Household members aged 15 years or over for each participating private dwelling In the 2013 wave, HILDA had 17,501 respondents 	Melbourne Institute of Applied Economic and Social Research		<ul style="list-style-type: none"> Melbourne Institute of Applied Economic and Social Research is funded to undertake primary analysis Data is available also to organisations and individuals with DSS-issued licenses 	<ul style="list-style-type: none"> All government entities can use longitudinal data to inform policy development, implementation and evaluation, and to highlight areas that may require further research
Longitudinal Study of Australian Children (LSAC)	A multidisciplinary study that analyses the impact of Australia's social and cultural environment on the development and wellbeing of children	Australian Government through DSS	The ABS undertakes multi-mode data collection every two years	<ul style="list-style-type: none"> Children (as well as parents, teachers and childcare workers) born between March 2003 to February 2004, and March 1999 to February 2000 In the 2014 wave, LSAC had 8,041 respondents 	Australian Institute of Family Studies	Conducted by the Australian Institute of Family Studies in collaboration with DSS, ABS, Department of Education and Training, Australian Children's Education and Care Quality Authority, and Department of Human Services	<ul style="list-style-type: none"> Australian Institute of Family Studies is funded to undertake primary analysis Data is available also to organisations and individuals with DSS-issued licenses 	<ul style="list-style-type: none"> Longitudinal analysis can also be used by non-government entities, academics, and researchers in the public and private sectors
Australian Longitudinal Study on Women's Health (ALSWH)	A cohort study that examines female physical and mental health, as well as psychosocial aspects of health (such as socio-demographic and lifestyle factors) and their use of health services	Australian Government through the Department of Health	ALSWH is a collaborative project conducted by the University of Newcastle and University of Queensland	<ul style="list-style-type: none"> Women who were aged 1-7, 18-23, 45-50 and 70-75 in 1996 Across its four cohorts, ALSWH has 21,925 respondents at most recent count 	University of Queensland and University of Newcastle	Department of Health (MBS/PBS/Aged Care), owners of state-based cancer, perinatal and admitted patients datasets in NSW, QLD, WA, SA, ACT and VIC (Cancer registry only)	<ul style="list-style-type: none"> University of Queensland and University of Newcastle are both funded to undertake primary analysis Data is also made available to approved, collaborating researchers 	

6.4 Australia's Core National Longitudinal Data Assets

There are dozens of longitudinal surveys in Australia and a near unquantifiable number of administrative and commercial datasets that have the potential to offer a longitudinal view. The Review has constructed a catalogue of 78 longitudinal datasets, composed of 43 ongoing surveys, 28 completed surveys and seven linked datasets (see Appendix D for a full list of the catalogued datasets). While not exhaustive, this catalogue offers an overview of the largest and most widely used longitudinal datasets in Australia.

In order to focus the analysis, the Review has sought to refine the catalogue by identifying Australia's initial 'Core National Longitudinal Data Assets'. This is a comparable concept to the ABS's Essential Statistical Assets, which aim to highlight datasets (longitudinal and other) that are critical to decision making in Australia. The Review has made the classification of Core National Longitudinal Data Assets by applying the criteria in the table to the right.

From this process, the Review has identified 15 longitudinal data assets that satisfy these criteria, composed of eight ongoing surveys and seven linked or 'curated' administrative datasets. These datasets should be considered the starting point of Australia's Core Longitudinal Data Assets. Over time, it is expected that these datasets will be preserved and strengthened, while other datasets that meet the above criteria will also be added. This should include longitudinal datasets held by states and territories, and other non-Commonwealth actors. Completed longitudinal assets could also be revived to become core longitudinal assets.

The initial Core National Longitudinal Data Assets identified by the Review are listed in the table overleaf.

Criteria	Description
Alignment	The data asset focuses on one or more ABS Essential Statistics Australia dimensions.
Priority	The data asset is currently (or likely in future) related to a national policy priority.
Usage	The data asset achieves high usage on national policy priorities.
National	The data asset provides national coverage. If sub-national, the data asset could be expanded (for survey data) or combined with other data assets (for administrative data) to provide national coverage or be usefully applied to a national context.
Unique	The data asset provides unique longitudinal insights into a national policy priority not provided by other data assets (longitudinal and other).
Linkable	The data asset contains requisite fields to support data linkages.
Contemporary	The data asset is current and well-maintained, or could it be made so with a reasonable effort or investment.
Secure for sharing	The data asset can be shared and accessed by approved users, either as is or in an adjusted state.
Permission	The data asset is permitted to be linked with other data assets.
Quality	The data asset meets standard industry quality thresholds.

CORE NATIONAL LONGITUDINAL DATA ASSETS

The table below shows the longitudinal surveys categorised by the Review as initial Core National Longitudinal Data Assets.

Core National Longitudinal Data Assets	Description
Surveys	
Australian Census Longitudinal 5% Dataset (2006 and 2011) (ACLD)	A dataset that brings together a 5% sample from the 2006 Census with records from the 2011 Census to create a research tool for exploring how Australian society is changing over time.
Household, Income and Labour Dynamics in Australia (HILDA) Survey	A household-panel study which began in 2001, comprising an Australia-wide sample. It involves annual interviews with all adult members for each participating household about economic and social well-being, along with health, labour market and family dynamics.
Growing up in Australia: the Longitudinal Study of Australian Children (LSAC)	A multi-disciplinary study which began in 2004 that analyses the impact of Australia's social and cultural environment on the development and wellbeing of two cohorts of children born in the late 1990s and early 2000s. The survey comprises biennial interviews about parenting, family relationships, childhood education and non-parental child care and health.
Longitudinal Surveys of Australian Youth (LSAY)	A study that began in 1995 of young people from age 15 as they move from school into further study, work and other destinations. LSAY involves a wide range of school and post-school topics including attitudes, aspirations and achievements related to school and work. The samples are drawn from students undertaking the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA).
Australian Longitudinal Study on Women's Health (ALSWH)	A cohort study that began in 1996 and examines the health of over 58,000 Australians across three cohorts of women above the age of 18. The study involves a triennial data collection that takes a comprehensive view of all aspects of health that assesses physical and emotional health; use of health services; health behaviours and risk factors; time use; socio-demographic factors; life stages and key events.
Ten to Men: The Longitudinal Study on Male Health (ALSMH)	A population based survey that began in 2011 and examines the health, across three cohorts, of males above the age of 10. The study aims to examine male health and its key determinants including social, economic, environmental and behavioural factors that affect the length and quality of life of Australian males.
Footprints in Time: the Longitudinal Study of Indigenous Children (LSIC)	A cohort study that began in 2008. Looking at two cohorts, it aims to improve the understanding of, and policy response to the diverse circumstances faced by Aboriginal and Torres Strait Islander children, their families and communities. It involves an annual survey about child physical and mental health, families, communities and access and usage of services.
Building a New Life in Australia – The Longitudinal Study of Humanitarian Migrants (BNLA)	A study that began in 2013 and analyses how humanitarian migrants settle into a new life in Australia. The study findings will identify factors that help or hinder the successful settlement of humanitarian migrants assist in improving policy development and program delivery.

CORE NATIONAL LONGITUDINAL DATA ASSETS

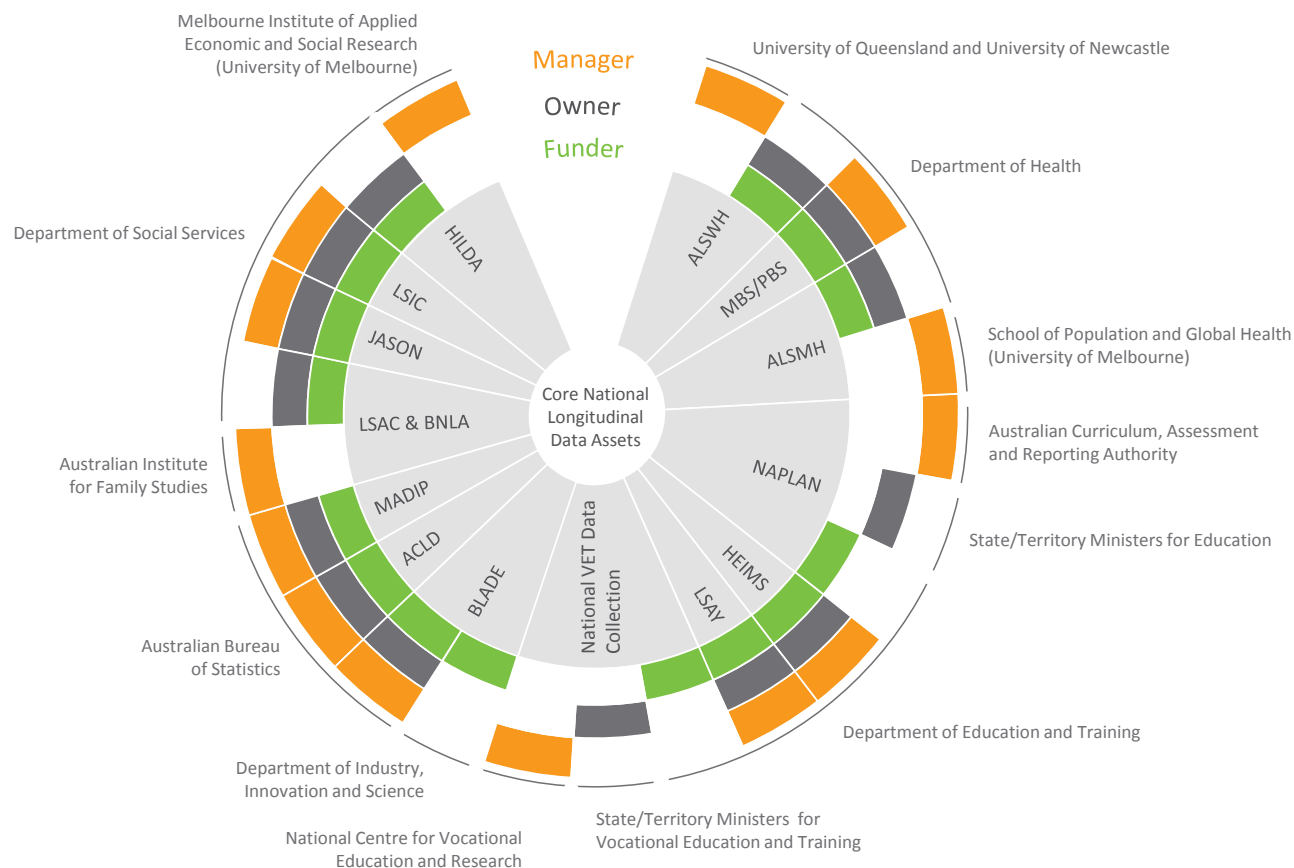
The table below shows the longitudinal administrative datasets categorised by the Review as initial Core National Longitudinal Data Assets. Note that these administrative data assets have been linked or curated in some way; they derive from various underlying raw administrative datasets.

Core National Longitudinal Data Assets	Description
Administrative	
Business Longitudinal Analytical Data Environment (BLADE)	The BLADE (formerly the Expanded Analytical Business Longitudinal Database) integrates administrative data from the ATO with collected survey data from the ABS for all active businesses in the Australian economy from 2001-02 to 2012-13.
The Longitudinal Dataset for the Investment Approach (JASON)	A longitudinal quarterly dataset drawn from a suite of episodic administrative datasets that includes information on individuals receiving: Pensions, Parenting Payments; Allowances, ABSTUDY; Family tax benefit, Child care payments and Parental leave pay; Supplementary payments and allowances; Concession cards.
Medicare Benefits Scheme/Pharmaceutical Benefits Scheme 10% dataset (MBS/PBS)	The MBS/PBS claims data are an administrative by-product of the Department of Human Services administration of its payment system for processing services and products that qualify for the Medicare or Pharmaceutical Benefits Scheme.
National Vocational Education and Training (VET) Data Collection	Datasets that include a comprehensive, integrated suite of statistical VET information. It contains data relating to VET activity, including traineeships and apprenticeships. The Unique Student Identifier initiative is currently being implemented to allow students to access records more easily and for the Australian Government to capture longitudinal data on VET student activity.
Multi-Agency Data Integration Project (MADIP)	An ongoing project to create an enduring, linked, publically accessible research dataset to help government agencies and researchers respond to nationally important policy and service delivery questions. Linked datasets are drawn from partner agencies that include DSS, Department of Health, DHS, Australian Taxation Office (ATO) and the ABS. MADIP is not currently longitudinal but is being designed to provide a longitudinal perspective.
National Assessment Program - Literacy and Numeracy (NAPLAN)	A dataset comprising the annual NAPLAN data for participating students in Years 3, 5, 7 and 9. Individual students can be tracked across years. The Australian Curriculum, Assessment and Reporting Authority (ACARA) is the independent statutory authority responsible for the overall management of the Australian National Assessment Program, in collaboration with representatives from all states and territories and non-government school sectors.
Higher Education Information Management System (HEIMS)	The national higher education data collection, encompassing data on students and higher education providers. The majority of individual students can be tracked across time between providers using the Commonwealth Higher Education Student Support Number (CHESSN).

GOVERNANCE MAP FOR THE CORE NATIONAL LONGITUDINAL DATA ASSETS

The diagram on the right depicts the various actors governing the Core National Longitudinal Data Assets. Whilst this presents a complex web of stakeholders, four Commonwealth entities help govern almost all of these assets:

- *DSS* is funder, owner and/or manager for five Core National Longitudinal Data Assets. It funds, owns and manages JASON and LSIC, and funds and owns but delegates the management of HILDA, LSAC and BNLA to research institutes.
- *Department of Education and Training (DET)* is funder, owner and/or manager for four Core National Longitudinal Data Assets. It funds, owns and manages LSAY and HEIMS, and it funds but does not own and manage NAPLAN and the National VET Data Collection (both of which are owned by the relevant State and Territory Ministers).
- *ABS* is part funder, owner and manager for three Core National Longitudinal Data Assets: the AGLD, MADIP, and the BLADE. However, MADIP involves linking data assets from other Commonwealth agencies including DSS, Department of Health, DHS, and the ATO. The BLADE involves linking ABS survey data with ATO data and information from the Department of Industry, Innovation and Science.
- *Department of Health* is funder, owner and/or manager for three Core National Longitudinal Data Assets. It is funder, owner and manager for the MBS/PBS 10% Sample Dataset. Whilst it funds and owns ALSWH and ALSMH, it delegates their management to tertiary institutions.



6.5 Content coverage and gaps

There are two main ways to analyse the coverage provided by Australia's Core National Longitudinal Data Assets. The first is to examine what they tell us about the 'life-course' from birth to death. The second is to examine what they tell us about various enduring policy themes, many of which cut across several stages of the life-course.

Both are considered below in relation to the coverage of the Core National Longitudinal Data Assets. Note that some of these gaps may be covered to varying degrees by other longitudinal assets outside of these initial Core National Longitudinal Data Assets identified by the Review.

Life-course coverage provided by the initial Core National Longitudinal Data Assets

One of the biggest benefits of longitudinal data is its capacity to tell us about individual change, transitions and trajectories. Longitudinal data can enhance our understanding of the Australian life-course, including by providing insights into key transition points in life such as youth to adulthood and working to retirement. For firms, this life-course may be much longer, but data on firm entry through all stages through to firm exit is important.

The Review has examined the coverage provided by the Core National Longitudinal Data Assets over seven stages of the life-course for individuals: child ready, school ready, life ready, relationship ready, work ready, retirement ready, and end of life ready.

The Core National Longitudinal Data Assets cover a spectrum of life stages. Three data assets cover every life stage (ACLD, HILDA and MADIP) whilst three cover

only one life stage (the BLADE, National VET Data Collection and HEIMS).

The most commonly covered life stage is the life ready stage, followed by work ready stage. The least commonly covered life stages are relationship-ready and end-of-life-ready, both of which are covered by seven data assets only. HILDA provides the broadest coverage, offering insights across the entire life-course.

This analysis changes in part when factoring in the depth (not just breadth) of coverage. Whilst life-ready and work-ready remain the top two in terms of breadth and depth of coverage, the school-ready stage has greater depth than the remaining Core National Longitudinal Data Assets. This is primarily due to targeted school-aged focus of LSAY, LSIC and NAPLAN participants.

Enduring policy theme coverage provided by the initial Core National Longitudinal Data Assets

While examining the life-course shows the sequential coverage of Australia's Core National Longitudinal Data Assets and highlights key transition points, we must also consider the coverage of cross-cutting policy themes that are not confined to any one stage of the life-course, such as mental health and disability.

To assist this analysis, the Review identified a broad set of enduring policy themes through a scan of more than 200 current Commonwealth, State and Territory Government priorities. These priorities were grouped by themes and filtered to identify those that are most amenable to longitudinal analysis (for example, infrastructure was excluded). The identification of these themes was also guided by the classifications in the ABS' Essential Statistics Australia.

This analysis shows that the policy themes that are subject to greatest breadth and depth of coverage by the Core National Longitudinal Data Assets are related to employment and schooling (primary and secondary). Primary care, higher education and housing and homelessness are the next most covered policy priorities. Crime and justice have the least amount of coverage.

The Review has itemised the main gaps in the coverage of the Core National Longitudinal Data Assets in Appendix E. These gaps have been synthesised from a combination of the Review's analysis and stakeholder consultations, and should be used as a starting point for further testing and analysis.

LIFE-COURSE AND POLICY COVERAGE OF THE INITIAL CORE NATIONAL LONGITUDINAL DATA ASSETS

		Life-course		Policy themes																																	
		Life-course stages							Health								Education & employment						Society & community														
		Pre-natal and child ready	School ready	Life ready	Relationship ready	Work ready	Retirement ready	End of life ready	Health and nutrition	Preventative health	Primary care	Hospitals	Chronic disease	Disability	Alcohol and substance abuse	Mental health	Aged care	Financing health care	Early childhood development	Schools	VET	Higher education	Employment	Industries and business	Skills and productivity	Social security	Family functioning	Domestic violence	Indigenous	Housing and homelessness	Community based services	Rural and regional development	Arts and culture	Youth	Ageing	Crime and justice	Immigration
Surveys	Australian Census Longitudinal 5% Dataset (2006 and 2011)																																				
	Household, Income and Labour Dynamics in Australia study																																				
	Longitudinal Study of Australian Children																																				
	Longitudinal Surveys of Australian Youth																																				
	Australian Longitudinal Study on Women’s Health																																				
	The Longitudinal Study on Male Health																																				
	Longitudinal Study of Indigenous Children																																				
	Building a New Life in Australia – Humanitarian Migrants																																				
Administrative	Business Longitudinal Analytical Data Environment																																				
	The Longitudinal Dataset for the Investment approach																																				
	Medicare Benefits Scheme/Pharmaceutical Benefits Scheme 10% Dataset																																				
	National VET Data Collection																																				
	Multi-Agency Data Integration Project																																				
	National Assessment Program – Literacy and Numeracy																																				
	Higher Education Information Management System																																				

Key: extent of coverage

		Broad and deep regular coverage.
		Broad coverage (with some depth), or narrow coverage with depth. Must be regular.
		Broad coverage and regular (but not deep); or deep but not broad and regular coverage.
		Some (or high level) coverage. May comprise irregular coverage in waves.
		No coverage.

7. Strengths of the longitudinal data system

Australia's longitudinal data offers strong insights into Australian life

Australia's existing longitudinal data holdings offer a wide range of insights across the Australian life-course and various public policy issues, as discussed above. While this coverage is by no means exhaustive, significant value is being derived from this longitudinal data coverage. It offers a strong mix of breadth across the life-course and depth into particular policy themes. HILDA provides a prominent example of this sort of quality in Australia's existing longitudinal data assets.

HILDA'S CONTRIBUTION TO OUR UNDERSTANDING OF AUSTRALIAN LIFE

HILDA is seen as a world leading household-panel study that is offering increasing value as time goes on. It has enhanced our understanding of many aspects of Australian life. This includes: how smoking bans have affected the health and smoking behaviour of Australians; the reasons for changes in Australia's birth rate; the personal and national costs of mental illness; the division of household chores between men and women; the impact of separation and divorce on income; and how health influences the ability to work.

Compared to other household panel surveys overseas, HILDA maintains exceptionally high re-interview rates, speaks with all (not just one) household members and covers an impressive breadth and depth of topic areas. Consequently, the findings from HILDA have gone on to be extensively cited by researchers from many fields and countries.

Key projects are unlocking the potential of administrative data

While there is not a strong overarching governance framework for Australia's longitudinal data, this diffuseness has enabled some nimbleness, competitiveness and innovation. In addition, there have been strong inroads into sharing information between governments to understand issues affecting Australians and deliver better outcomes.

For example, there are various initiatives currently underway to link administrative data, such as the MADIP and JASON. The NCLD was formed to increase coordination in some longitudinal surveys. The Deputy Secretaries Data Group is implementing a roadmap of various programs and reforms related to public sector data management,

and Data61 is using data analytics to connect disparate government datasets and publicly release them through open data platforms.

There are also pockets of excellence at the state level and in the non-government sector. For example, a significant data linkage and longitudinal data infrastructure has been created in Western Australia. There are numerous other examples of researchers working with governments to use longitudinal data to provide deep insights on the operation and success of government programs. All of these initiatives are critically enabled by rapid improvements in technology, which is providing greater options to protect privacy, link datasets and share information amongst trusted networks.

DATA LINKAGE IN WESTERN AUSTRALIA

Data Linkage WA was established as a collaboration between the Department of Health Western Australia, University of Western Australia, Telethon Kids Institute and Curtin University. It manages a system which links WA's core population health datasets. These linkages are created and maintained using rigorous, internationally accepted privacy preserving protocols, probabilistic matching and extensive clerical review. Health data can be requested for ethically approved research, planning and evaluation projects.

Longitudinal data has delivered strong privacy protections

The Review is not aware of a privacy breach in the history of Australia's longitudinal data. This is a significant testament to both the privacy frameworks and the culture of vigilance and conscientiousness of the main actors in the longitudinal data system.

This is a need for constant care in the protection of privacy and management of the 'social licence' for longitudinal data. Community concerns over the recent wave of the Census of Population and Housing underscore the importance of these efforts.

Several stakeholders commented to the Review that this social licence is well managed in respect of specific longitudinal surveys, where participants often have a long and proud association with a study. The challenge is to build this confidence and trust for administrative longitudinal data and linkages. These issues will be canvassed further in the Productivity Commission's Inquiry into Data Availability and Use.

8. Weaknesses in the longitudinal data system

SYSTEM WEAKNESSES

While Australia has made some strong investments in longitudinal data, the Review has identified some weaknesses that are hampering its potential to enhance our understanding of Australian life and strengthen the policy making process. Many of these weaknesses share parallels with the challenges identified in the Public Sector Data Management Report .

1. The potential value of Australia's longitudinal data is not being realised

2. Australia's planning for and investment in longitudinal data is uncoordinated and short term

3. Longitudinal data is often collected, documented, stored and accessed inconsistently

4. It is often difficult for researchers and policymakers to access and analyse longitudinal data

5. There is insufficient linking of longitudinal survey and administrative data

6. There is insufficient capability to analyse longitudinal data and translate its insights

1. The potential value of Australia's longitudinal data is not being realised

There is a cadre of public policy professionals, researchers, and data analysts who extol the capacity of longitudinal data to enrich, and even change, our understanding of Australian life. But outside of these specialised circles, the value of longitudinal data tends to be less well understood.

This is compounded by the dispersed nature of Australia's longitudinal data collections. There is no one place for a data analyst to go to find and analyse longitudinal data. Even once a relevant longitudinal data asset has been located, access can be cumbersome. It may also be many years before the data can provide useful insights. None of this suits the fast-paced reality of responding to public policy challenges.

The under-appreciation and under-promotion of the value of longitudinal data has three main effects:

- Longitudinal data collection and analysis does not tend to be commissioned in a way that would enable it to best inform current and future public policy challenges. Instead, many longitudinal data studies, particularly when undertaken for academic purposes, fail to be translated into the sort of policy insights sought by decision makers.
- Longitudinal analysis of surveys and administrative data tends to be under-utilised in the formulation, implementation and evaluation of public policy in Australia. Instead the search for causation too often relies on cross-sectional data, qualitative evidence and anecdotes.
- Funding for longitudinal studies tends to be short-term and vulnerable. Until the value of longitudinal data is fully appreciated, it will inevitably slide down the list of governments' many competing priorities.

Improving policy makers' understanding of the value of longitudinal data would help to stimulate the demand for longitudinal analysis. This requires a more effective promotion of the value of longitudinal data and its capacity to enhance the public policy making process. It also requires measures to better track the use and impact of longitudinal data on policymaking and academic research. In turn, this would help to solidify support for investing in longitudinal data as a national asset both in the government and non-government sectors.

The first step might be for longitudinal data to have a clearer role in the implementation of the Public Sector Data Management Report. Longitudinal data should be conceived as a central part of the Government's approach to improving the use of data and much longitudinal data should be considered 'High Value Datasets'.

2. Australia's planning for and investment in longitudinal data is uncoordinated and short term

Investing in the production of longitudinal data is often a lengthy, complicated and costly process. It requires careful planning to ensure that the data collected in surveys or extracted from administrative datasets will make the highest value contribution to our understanding of Australian life, while protecting privacy.

There is currently no central point of advice for the Australian Government on Australia's longitudinal data needs, nor is there a mechanism to coordinate the Commonwealth's longitudinal investments with those made by the states and territories.

Instead planning tends to be reactive to funding proposals for specific longitudinal surveys made by various departments and stakeholders, often without sufficient regard to how these investments fit into the broader longitudinal data landscape. Relatedly, there is no longitudinal administrative data framework that can be followed by Commonwealth agencies to develop and make their administrative data available for longitudinal analysis by both internal and external data analysts.

This situation creates the risk of commissioning new surveys that duplicate existing surveys or administrative datasets, as well as failing to prioritise gaps that should be filled in Australia's longitudinal data coverage. The lack of a coordinated funding regime also creates unevenness in Australia's longitudinal asset investment, both in terms of funding levels and length.

The Australian Government has recently taken some steps to increase the coordination of its longitudinal surveys and broader approach to data, including through:

- the endorsement of a set of principles to govern integration of Commonwealth data for statistical and research purposes, along with the establishment of the Cross Portfolio Data Integration Oversight Board
- the establishment of the NCLD to bring together the longitudinal surveys managed by DSS
- the work of the Secretaries Data Group and Deputy Secretaries Data Group to coordinate various data-related projects across the Commonwealth
- MADIP, a five agency partnership, which was set up to create an enduring, linked, publicly accessible research dataset encompassing both surveys and administrative data, in its maturity.

The Australian Government may want to build on these measures through a special focus on coordinating longitudinal data investments, both within the Commonwealth and across jurisdictions. This could include a central place for the Australian Government to receive advice on Australia's longitudinal data needs and the various options to meet them. The Review's recommended approach is discussed in the next part of this report.

3. Longitudinal data is often collected, documented, stored and accessed inconsistently

Practices in the production of longitudinal survey data vary widely. This inconsistency is found across most parts of the longitudinal data production value chain:

- **Collection:** There is a lack of widely accepted and consistently applied standards related to the creation of longitudinal survey data. This results in a wide variety of survey designs, questions and collection methods. This also includes the absence of strong protocols for processing and preparing data in accordance with privacy and confidentiality requirements.
- **Documentation and metadata:** There are no standards for data documentation and associated metadata. As a consequence a wide range of approaches are taken. These issues encompass all metadata, including documentation of data collection methods, conditions or permissions given by survey respondents, survey questionnaires and data elements. These issues apply to both surveys and administrative data, with many administrative datasets having limited or no documentation at all (noting that little administrative data is publicly accessible).
- **Storage:** In cases where data files are provided, data analysts may be required to meet high standards for storage and security so as to minimise the risk of unauthorised individuals accessing data and breaching confidentiality and privacy. At times, these standards may be unnecessarily stringent, difficult and costly to meet, and may be inconsistent between datasets.
- **Evaluation:** With some notable exceptions, such as the Australian Longitudinal Study on Women's Health and the Australian Longitudinal Study on Men's Health, few datasets are subject to consistent evaluation. Every dataset should be regularly evaluated to ensure it is providing appropriate value and influence. Each dataset should seek to have optimal efficiency, accessibility and usage.

4. It is often difficult for researchers and policymakers to access and analyse longitudinal data

Stakeholders frequently nominated the difficulties associated with accessing longitudinal data as one of the system's biggest weaknesses. There is no 'front door' to longitudinal data assets, nor is there a comprehensive catalogue of current longitudinal data assets.

The application process for analysts to obtain access to longitudinal data varies greatly, depending upon the dataset and whether formal access arrangements have been established. In the case of surveys such as HILDA, the application process is relatively straightforward and quick. However, in other cases potential users face long waits to understand what data exists, how it will meet their needs (if at all), and how it can be accessed and analysed. This diminishes the overall transparency and accountability of the system.

A particular obstacle is that there are few curated administrative datasets that have been prepared for access and analysis by approved researchers. Researchers commented to the Review that they feel "locked out" of accessing administrative data. Even once a user is permitted access to data, the need to preserve privacy and confidentiality often means that restrictive access environments are imposed. The lack of proper documentation and metadata for many administrative datasets also hinders their usability.

ABS' REMOTE ACCESS DATA LABORATORY

The ABS's Remote Access Data Laboratory allows survey data users to submit queries in analytical languages against confidentialised unit record files that are kept within the ABS environment. The results of the queries are then checked before they are made available to the users. The ABS does offer more expansive access through the ABS Data Laboratory but such access is limited to a secure location within ABS premises.

4. It is often difficult for researchers and policymakers to access and analyse longitudinal data (continued)

Where researchers do access, document and clean the data, there is no process for ensuring that the information (such as the code used for cleaning) is shared with custodians and other researchers using the data source.

Such processes can be a disincentive for researchers and policymakers, both domestically and internationally, to seek access to Australia's longitudinal data assets. This depresses demand for Australia's longitudinal data and forgoes great potential to generate new insights on Australian life.

Additionally, there are many 'single use' datasets that researchers are required to destroy after use. While this is done in the interests of confidentiality and privacy, it does not take into consideration the considerable resources that go into cleaning datasets and possible future applications of the data (including potential future linkages).

5. There is insufficient linking of longitudinal survey and administrative data

Linking data (as best enabled by a common, unique identifier) has the potential to multiply the value of longitudinal data. Administrative data can be used to supplement and validate longitudinal surveys, and in some cases diminish the burden on respondents and decrease survey costs. It can also provide additional data points to help fill in the picture between survey waves, and help researchers to stay in touch with participants to reduce attrition. While administrative data has the advantage of coverage of the population it relates to, it lacks detail that only survey data can provide. Thus, it is the *interaction* of longitudinal survey and administrative data that holds the greatest potential to enlarge our understanding of Australian life.

Australia is making some progress towards increasing data linkages:

- **Linking administrative data:** There are various projects underway to link administrative datasets. For example, the MADIP has commenced its first phase by bringing together a concordance of linkage keys for the following administrative datasets: social security and related information, Medicare enrolments and claims information, personal income tax information, and Census data. Future phases may expand to include a variety of other datasets and could potentially be linked with longitudinal surveys.

Another leading example is JASON – a longitudinal quarterly dataset based on a suite of episodic administrative datasets extracted from the DHS Enterprise data warehouse. Information consolidated in the database include: Pensions (Age, Carer, Disability Support Pension and others); Parenting Payments; Allowances (Newstart Allowance, Youth Allowance, Austudy and others); ABSTUDY; Family Tax Benefit, Child Care Payments and Parental Leave Pay; Supplementary Payments and Allowances; Concession Cards. This data could also be potentially linked with longitudinal survey data.

The BLADE also integrates administrative data from the ATO with ABS collected survey data from all active businesses in the Australian economy from 2001-02 to 2012-13. This is creating an enduring firm level statistical asset that will increase the capacity of the researchers to undertake firm-level analysis of micro-economic drivers of performance, competitiveness and productivity.

5. There is insufficient linking of longitudinal survey and administrative data (continued)

- **Linking longitudinal survey and administrative data:** Some longitudinal surveys have established substantial links with administrative data. The best example in Australia is LSAC, which links the survey responses to a number of administrative databases, including Medicare data, NAPLAN results, National Childcare Accreditation Council data, and Census data. However, these linkages involve a difficult process of obtaining the consent of participants, obtaining the agreement of the data holders to permit access to the data, and conducting probabilistic matching to connect the administrative data to the survey participants.
- **Linking survey data:** The ACLD links a 5 per cent random sample from the 2006 Census with the 2011 Census using data linkage techniques without name and address. It is envisaged that the 2016 Census and subsequent Censuses will be added in the future, as well as administrative datasets.

Despite this progress, there are several barriers that still impede greater linkages between longitudinal survey and administrative data:

- **Consent:** The written consent of longitudinal survey participants is required before their administrative data can be accessed. This process can be cumbersome and consume valuable interview time. It also creates a tension between maximising consent and fulfilling the legal and ethical requirement for participants to provide fully informed and continuous consent.

Non-consenters or participants who subsequently withdraw their consent create the risk of non-response bias that can skew the data. However, the experience of data collectors in Australia and overseas is that these consent rates tend to hover above 90%, depending on the study, cohort and type of administrative data. The default mode of consent required (including opt-in vs opt-out arrangements) can greatly impact these consent rates.

- **Regulation:** The Privacy Act and the associated Australian Privacy Principles provide the primary framework for managing personal information. With the exception of Western Australia and South Australia, the Australian states and territories have their own privacy legislation. There are also separate bodies of legislation governing the collection, use and disclosure of some specific types of information. Additionally, there are various agency-specific legislation, policies and procedures that govern the collection, use and disclosure of information held within administrative datasets.

This complex landscape has been seen by some stakeholders as an inhibitor to realising the potential of longitudinal data. It tends to cause confusion for data holders and engenders a cautious approach to sharing and linking data. As the Public Sector Data Management Report commented, “privacy concerns and cautious interpretation of legislation are holding the Australian Public Service back from making the most of its data.”

For example, the distinction between de-identified and confidentialised data is not clear and is difficult to apply in practice. This has implications for how privacy regulations should apply. Confidentialised data, and even aggregate and synthetic data, may have the potential to be converted into identifiable data, especially when combined with other data. Techniques are available to assess the probability of re-identification, but a risk-based approach is not about ensuring this probability is zero, but rather assessing the probability of misuse of the information. This depends on not just safe data, but also safe people, safe output, safe project and safe setting (known as the ‘five safes’). In addition, the cost of any re-identification should be taken into account. These issues are being explored further in the Productivity Commission’s Inquiry into Data Availability and Use.

5. There is insufficient linking of longitudinal survey and administrative data (continued)

- **Culture:** A strong theme from the Review's consultations was that there is a risk-averse culture to sharing data across government, both within the Commonwealth and across jurisdictions. In many cases these cultural factors present a higher barrier to data linkage than any regulatory restrictions.

Administrative data holders have a number of concerns that inhibits their willingness to share data, including that:

- their data may not be well maintained, such that sharing it will expose them to external scrutiny
- sharing their data may result in adverse findings for their Department and Minister
- sharing or integrating their data will dilute their access to and control of the data
- they do not have adequate resources and capacity to properly engage in data sharing initiatives.

Consequently, sharing often depends on trust between various data asset owners that personal information will be protected and the quality and type of analysis will be appropriate. As noted in the Public Sector Data Management Report, this has created an environment where:

- it can take several years and multiple 'memorandums of understanding' to establish data sharing arrangements between government agencies
- when projects occur, often only a few tables are produced from important linked data, which is then destroyed
- sharing data with states and territories is burdensome
- the research community is frustrated by the red tape that impedes greater data sharing, the time it takes to be granted access to data and the inconsistency in user charging for data.

- **Matching:** It is not always possible to match survey participants to administrative data. Absent a unique identifier, researchers need to employ probabilistic matching techniques using identifying variables, but this has its limitations. It also comes at a cost that can outweigh any efficiency derived from eliminating survey questions. Ensuring that data is linked in an enduring manner has the potential to reduce costs in the future. This is potentially best realised through a centrally-held catalogue of linked data assets needs for government and non-government researchers to access and understand potential linkages.

THE POTENTIAL VALUE OF A UNIQUE STUDENT IDENTIFIER

The Review heard from many stakeholders that linking school student data presents one of the biggest opportunities for strengthening longitudinal insights on school education and student achievement. A unique student identifier could be used to link data across the life cycle of learning, from early childhood to post-school pathways.

The Productivity Commission Inquiry into the National Evidence Base for School and Early Children has invited consideration of the cost and benefits of expanding the Unique Student Identifier scheme (which allows exact matching of VET information to identify records in different datasets that refer to the same person) to the schooling and early education sectors. The Inquiry is scheduled to be completed by end of 2016.

6. There is insufficient capability to analyse longitudinal data and translate its insights

A perfect system for commissioning, collecting, linking and accessing longitudinal data would achieve little without the capability to analyse the data and draw insights from it. There is a global under-supply of data and analytics skills. Australia and the Australian Public Service (APS) are no different. While there are pockets of advanced data analytics skills across the APS, most agencies seem to be suffering from a shortfall. As the Public Sector Data Management Report noted, “capability to use data is patchy across the APS”.

Longitudinal data analysis requires an especially high skill level. One of the biggest promises of longitudinal data is that it can move beyond the correlations shown by cross-sectional data and start to point towards causation. This often requires the application of sophisticated statistical models and the use of specialised statistical programs. Such skills are typically developed in the third year of university studies in statistics or econometrics, after first mastering cross-sectional analysis.

This technical capability needs to be complemented with the ability to translate longitudinal analysis into findings and advice that can be consumed by decision makers. Several stakeholders commented that there are too few examples of insights from longitudinal studies being presented in policy-relevant formats for non-expert users. Instead much analysis tends to languish in academic papers and dense technical reports.

The Public Sector Data Management Report committed to building the data and analytics capability across the APS by bolstering existing efforts, partnering externally and investing in pockets of excellence. This agenda is being overseen by the Deputy Secretaries Data Group. Progress has been made on several fronts:

- The Australian Public Service Commission (APSC) is leading a collaborative effort to develop a strategy with government, industry, and academia to build data and analytics capability. This includes the creation of a ‘Data Skills and Capability in the APS’ framework and an APS Data Literacy program.

- Australian National University, in partnership with DHS is providing a postgraduate (Graduate Diploma and Masters) data skills course.
- Data61 is working with the Deputy Secretary Data Group to propose a Data61 Academy.
- A whole-of-government Data and Analytics Centre of Excellence has been established in the ATO to enable a common capability framework for analytics, as well as an opportunity to share technical knowledge, skills and tools. The Centre will also help build collaborative arrangements with tertiary institutions to aid the development of analytics professionals.
- Data champions have been established to promote the use, sharing and re-use of data within their organisations and across the APS.

There is a need to ensure that these initiatives deliver the sort of specialised skills required to get the most out of longitudinal data. This might necessitate the creation of initiatives specifically focussed on longitudinal analysis capabilities.

There is also a need to focus on sustaining and rewarding these sorts of skills. In order to progress in the APS, specialist skills are often overlooked in favour of generalist and managerial skills. This creates the risk of losing longitudinal data expertise to the private sector.

9. Opportunities for the longitudinal data system

A consistent theme throughout the Review's consultations was that this is the right time for a 'step change' to be made to Australia's longitudinal data system and assets. The main opportunities in the current environment are set out below.

The public data agenda has momentum

There is growing demand for quantitative analysis to support evidence-based policy in Australia. There is also a growing recognition of the value of public data as a national asset. The Australian Government is taking measures to realise these opportunities, such as through the work of Public Sector Data Management Report, the Public Data Policy Statement and various 'high value projects' that are seeking to improve data access, use and innovation. As noted above, there are also various cross-agency projects underway to increase data linkages and prepare curated administrative datasets.

The Government's commitment to adopting an 'Investment Approach' to welfare reform presents a particular opportunity for longitudinal data. Longitudinal survey and administrative data is vital for the Government to be able to identify people most likely to remain on income support for long periods of time and design early interventions to prevent dependence.

The longitudinal data system needs to contribute to this momentum and play a key role in helping the Government to realise the full potential of public data. The challenge is to make it easy for users to access and analyse longitudinal data, especially when drawing on multiple datasets. This increases the importance of having consistent data standards, application and access requirements, and analysis tools.

Technology is creating new possibilities

There are rapid developments in information technology that increase the capacity for governments to create and analyse longitudinal data, in ways that preserve confidentiality and privacy. For example, the Australian Government is a holder of very large administrative datasets, such as the Centrelink and Medicare databases, that have been difficult to analyse in the past due to their large size. 'Big data' tools and techniques mean that it is now possible to intensively analyse these datasets to generate insights.

Additionally, data can now be made available for data analysis in ways that preserves privacy and confidentiality with increasing ease. Tools are available that enable data analysis to be undertaken in such a way that data users cannot access 'identifying' information. This has the potential to greatly expand access to data and increase its application to the policy making process.

We can learn from international approaches

Australia can learn a great deal from longitudinal data initiatives both within Australia and overseas. Many of the issues being confronted in the development of Australia's longitudinal data system are not unique in a global context. Countries such as New Zealand and Canada have embarked upon wide ranging activities to collect and disseminate longitudinal data. Similarly, Australia's state and territory governments are also seeking to bring together their data collections to better inform policy development – there is much potential for collaboration.

By way of example, there is a degree of caution among Australian state and territory governments to share and link student data encompassing early childhood and schooling. In the United States, concerns by state and local governments have been addressed through the provision of federal government financial support for the development of longitudinal data systems that enable students to be tracked over time, including into post-school education and training.

Australia should draw on these sorts of approaches in formulating its own solutions to such issues. An overview of some of these international experiences is provided in Appendix F.

Similarly, Australia has much to contribute to longitudinal data best practice through the cultivation of its own datasets and development of its longitudinal data system. As discussed above, longitudinal surveys such as HILDA are considered world-class and generate much attention, both locally and globally. There is also the potential to lead the way on developing administrative data linkages, strengthening access arrangements and developing platforms for longitudinal analysis.



PART III THE PATH FORWARD FOR AUSTRALIA'S LONGITUDINAL DATA SYSTEM

10. The vision for the future longitudinal data system

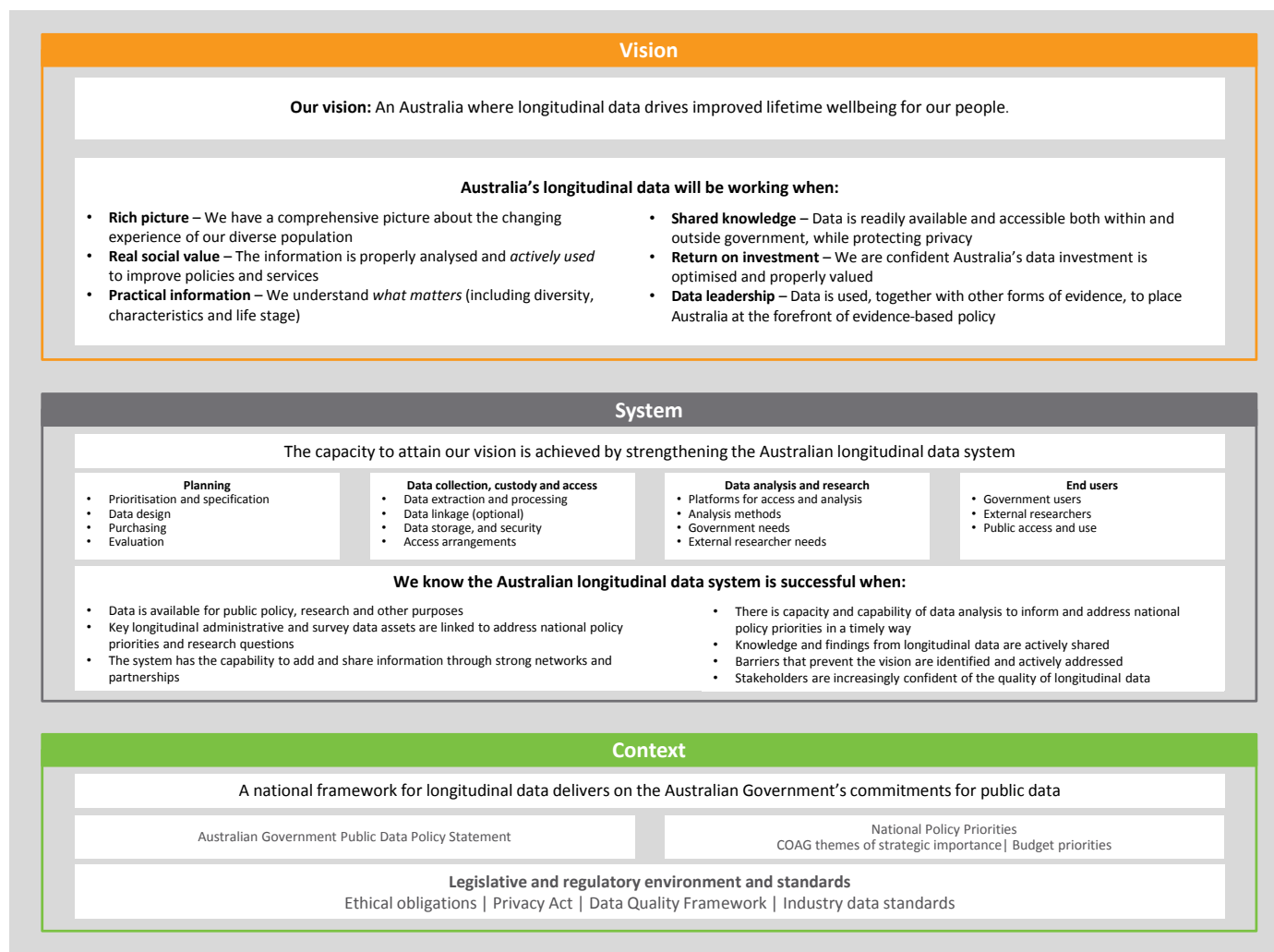
The Review has found that longitudinal data has the power to enrich our understanding of Australian life in a way that other data cannot. The Review is seeking to unlock the potential of this data by optimising the longitudinal data system and the Government's investment in it.

To this end, the Steering Committee has set a vision for Australia's future longitudinal data system whereby longitudinal data drives improved lifetime wellbeing for Australians. Australia's longitudinal data system should be world-class in providing a rich picture of our population and businesses. It should deliver timely, practical and relevant information to improve policies and services to citizens – reaching across enduring policy themes such as health, education and employment, and society and community.

This vision will be achieved by fostering a system that works effectively across the longitudinal data value chain, from planning the development of longitudinal datasets to the policy application of the insights generated. This system needs to be closely aligned with the broader public data management agenda, as well as initiatives in the states and territories, and non-government sectors.

This vision, its context and the connection to the longitudinal data system are presented in the diagram to the right.

VISION FOR AUSTRALIA'S FUTURE LONGITUDINAL DATA SYSTEM



11. Recommended strategy to realise the vision for the longitudinal data system

Achieving the vision requires a strategy and practical actions to improve the longitudinal data system. The Review's recommended priorities to achieve the vision are centred on the themes of preserve, strengthen and invest. An overview of these recommended priorities is set out below.

1. Preserve existing longitudinal data assets with sufficient funding to realise their full value

The Australian Government has made significant investments in longitudinal data over a sustained period of time. This provides a strong base of evidence on the life-course and various public policy issues. The first priority should be to sustain and capitalise on this evidence-base to ensure that the long-term value of the existing longitudinal data assets is realised. This requires a coordinated approach to the development of a funding model which can sustain and optimise longitudinal data, particularly for the Core National Longitudinal Surveys which are in need of ongoing funding. This may include exploring co-funding mechanisms and partnerships, as well as options to improve the efficiency and operation of surveys.

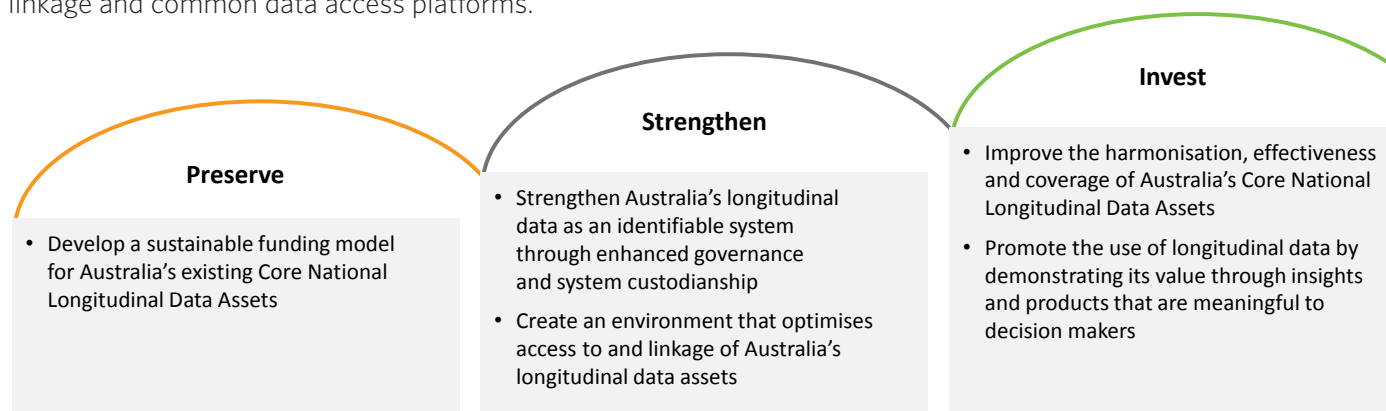
2. Strengthen the longitudinal data system

One of the Review's key findings is that Australia's current longitudinal data system is fragmented and uncoordinated. There is a need to strengthen this system through enhanced governance and coordination. The Review recommends that this be achieved through a federated governance model, whereby longitudinal data asset owners continue to manage their assets but are supported by a Longitudinal Data System Custodian and an Advisory Council. These bodies would help to facilitate and promote a consistent approach to longitudinal data planning, investments and standards on an opt-in basis.

There is also a need to foster an environment that optimises access to and linkage of Australia's longitudinal data assets. This requires addressing various barriers related to privacy, best practice standards for metadata, data access protocols, data linkage and common data access platforms.

3. Invest in the harmonisation, coverage and impact of longitudinal data

Once the longitudinal data system has been strengthened, the priority should be to expand the coverage, depth and impact of Australia's longitudinal data assets. This may include considering options to fill gaps in the life-course coverage of Australia's longitudinal data, such as the early years, Indigenous perspectives, interpersonal violence, and retirement and ageing. There should also be a focus on investing in the promotion and application of longitudinal data through projects that demonstrate its value and potential application to the decision-making process.



Priority 1. Preserve existing longitudinal data assets with sufficient funding to realise their full value

Action 1.1 Develop a sustainable funding model for Australia's Core National Longitudinal Data Assets

There is significant value in Australia's existing longitudinal data assets, which will increase over time. These assets should be viewed as strategic national resources that are of high public value. As public goods, their value extends beyond any one portfolio, department or level of government. They enhance policymaking and academic research across a wide range of domains.

Accordingly, the immediate priority should be to preserve and unlock the value of these assets, starting with the initial 15 Core National Longitudinal Data Assets identified by the Review, with a particular focus on the longitudinal surveys.

Develop a sustainable funding model for longitudinal surveys

There is a need to develop a long-term and sustainable funding model for longitudinal surveys. Most of the surveys identified by the Review as Core National Longitudinal Data Assets are funded through departmental expenditure and have been subject to recurring efficiency dividends. Over the same period, the labour costs required to administer surveys have risen. Some of these surveys are being sustained through an interim funding arrangement that is due to expire in June 2017, and funding for others will cease in 2019.

In the absence of a sustainable funding model, these surveys may need to be reduced in scope or quality, and much of the potential of Australia's longitudinal data evidence base could be lost. Australia should learn from the Canadian experience on the importance of preserving investments in longitudinal surveys.

LEARNING FROM CANADA'S EXPERIENCE WITH LONGITUDINAL SURVEYS

Canada's recent approach to longitudinal data underscores the importance of sustaining investments in longitudinal surveys. Statistics Canada has ceased a number of longitudinal studies (or longitudinal elements of broader studies) in the past twenty years, including the Survey of Labour and Income Dynamics (SLID) and the National Longitudinal Survey of Children and Youth.

As a result, Statistics Canada relies heavily on administrative data to provide a longitudinal perspective. It maintains only a handful of longitudinal surveys (such as the Longitudinal and International Study of Adults). This has led to significant gaps in Canada's longitudinal knowledge base, particularly in the areas of early childhood, ageing and employment. For example, the SLID is now a cross-sectional survey that provides data on aggregate poverty rates but it no longer provides individual change, transitional and causal insights.

Furthermore, Statistics Canada lacks critical experiential information (such as behavioural motivators) which can only be provided by longitudinal surveys. The Review's consultations with Statistics Canada suggest that, given the opportunity again, it would have sought to reduce the cost and respondent burden of its longitudinal surveys, rather than ceasing them altogether.

The need to renew funding for several longitudinal surveys presents an opportunity to move towards a more stable and better coordinated funding arrangement. This should occur through the development of a joint NPP from the owners of the Core National Longitudinal Data Surveys – on an opt-in basis – for a new round of funding. This proposal should also give consideration to the option of shifting funding for the surveys from departmental to administered expenditure. This would enable expenditure on the surveys to be more clearly defined, and separated from departmental efficiency dividends.

Priority 1. Preserve existing longitudinal data assets with sufficient funding to realise their full value (continued)

The development of this joint proposal should also canvass opportunities to better coordinate and optimise the surveys to demonstrate efficiency. Currently, the eight Core National Longitudinal Surveys are funded by four different Australian Government agencies, and managed by seven different organisations. There may be scope for improving the efficiency of survey administration and funding.

Over the medium term, the owners of the surveys should also strive for a co-funding or consortium model comprised of contributions from partners in other jurisdictions and the research sector. For example, there may be an opportunity to partner with universities or research institutes to utilise funding as part of the National Collaborative Research Infrastructure Strategy (NCRIS).

Develop a sustainable funding model for longitudinal administrative data

Much of the longitudinal administrative data identified by the Review as Core National Longitudinal Data Assets is currently being funded on a project-by-project basis, often with co-contributions from partnering agencies. Where this approach is working and sustainable, these arrangements should continue without interference. Other longitudinal data administrative linkage projects may require a longer-term and more coordinated model. A funding model for these administrative data linkage projects should be developed through a joint NPP from the owners of the Core National Longitudinal Data Assets, on an opt-in basis. This should be developed in the implementation phase of the Review.

EXAMPLES OF CONSORTIUM APPROACHES TO DATA INITIATIVES

SA NT DataLink provides a linkage service to enable research and policy analysis of de-identified data from multiple databases. It operates as a consortium of South Australian and Northern Territory government representatives, universities and research institutes, and non-profit organisations. Consortium members provide substantial funding and in-kind contributions for the operation of the SA NT DataLink. In addition, the Australian Government has provided financial support through NCRIS and the Population Health Research Network.

The Sax Institute is a national leader in promoting the use of research evidence in health policy, and aims to be the bridge between researchers and health decision makers. It receives core funding from the NSW Ministry of Health and is supported by other government, non-government, philanthropic and competitive research funding agencies, including the National Health and Medical Research Council, the Australian Government Department of Education and Training, and Cancer Council NSW.

Priority 2. Strengthen the longitudinal data system

Action 2.1 Strengthen Australia's longitudinal data as an identifiable system through enhanced governance and system custodianship

The public data landscape is complex and rapidly changing. While the custodians of longitudinal data have realised significant achievements operating separately, the Review has found that there is an opportunity to strengthen the longitudinal data system through increased coordination. This approach aligns with the Australian Government Public Data Policy Statement's emphasis on building partnerships and increasing engagement across the public data system.

The Review considered various potential governance models to overcome these concerns about fragmentation and lack of coordination. These potential models broadly sit along a spectrum of decentralisation and centralisation.

Australia's current approach to longitudinal data sits towards the decentralised end of the spectrum. Longitudinal data investments are sought separately for each data asset and distributed by various Commonwealth and state agencies, universities and research institutes. Data planning and collection is undertaken by a wide range of creators and custodians, often with different approaches to data design and documentation. Data is stored separately by these asset owners, with numerous approaches to data storage and access. Data linkage tends to occur on a project by project basis with the assistance of authorised linkage authorities and cross-agency working groups. The resulting data is then analysed by a wide range of users.

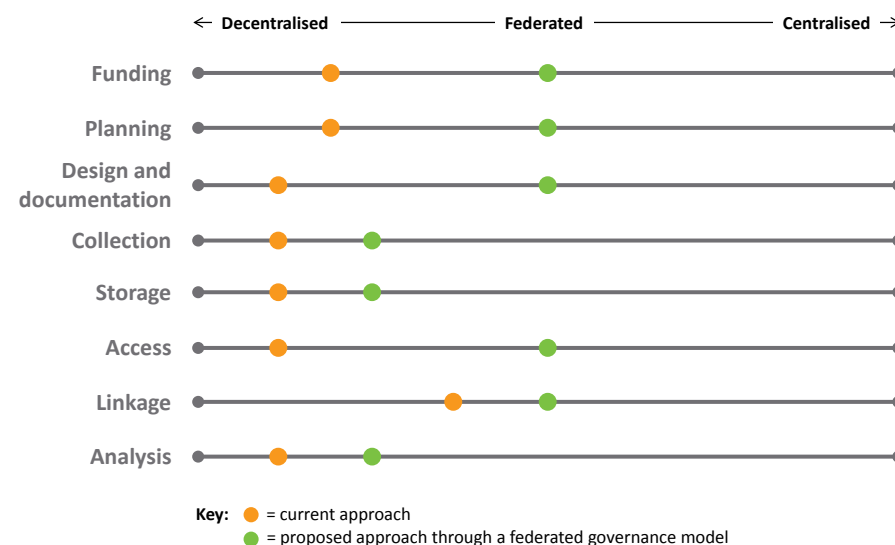
The Review does not propose a drastic shift towards centralisation. Trying to compact all longitudinal data assets and projects into a centralised system would risk stifling flexibility and innovation.

Instead, the Review recommends a federated governance structure under which:

- existing Longitudinal Data Asset Custodians would continue to have day-to-day management of their data assets, including responsibility for strategic planning and funding proposals
- a Longitudinal Data System Custodian would help to support data asset custodians by facilitating and promoting more consistency and cooperation
- a Longitudinal Data System Advisory Council would oversee the work of the System Custodian, set planning priorities and provide coordinated advice to government on Australia's future longitudinal data needs.

This recommended federated governance model is depicted along a decentralised/centralised spectrum across key domains in the figure below. Note that these shifts that would occur over time, in accordance with a feasible implementation schedule.

SPECTRUM OF GOVERNANCE APPROACHES



Priority 2. Strengthen the longitudinal data system (continued)

LONGITUDINAL DATA ASSET CUSTODIANS

Roles and responsibilities

Longitudinal Data Asset Custodians would continue to be responsible for the day-to-day management and custodianship of their data assets. The pre-existing governance arrangements for each of these data assets would remain in place.

The Longitudinal Data Asset Custodians would work alongside the Longitudinal Data System Custodian and Advisory Council as support mechanisms to help facilitate increased consistency and cooperation across the longitudinal data system. For example, the Longitudinal Data Asset Custodians would be able to choose to opt-in to best practice standards coordinated by the Longitudinal Data System Custodian.

The Longitudinal Data Asset Custodians would also be responsible for developing strategic plans for their data assets, encompassing issues such as data planning, documentation, data access arrangements, data linkages arrangements, and funding proposals. The Longitudinal Data System Custodian would help to coordinate these strategic plans, which would then be submitted to the Longitudinal Data System Advisory Council to enable it to provide informed advice to Government on Australia's future longitudinal data needs.

Membership

The Longitudinal Data System Custodian and Advisory Council should be seen as resources that are open to all actors in the longitudinal data system. However, the Review envisages that the initial focus of the system will be on the funders and custodians of the Core National Longitudinal Data Assets. This encompasses DSS, Department of Health, DET, ABS, and Department of Industry, Innovation and Science. It should also include the various custodians who have day-to-day management of these data assets. Once these governance arrangements have been established, the Review expects that this initial focus will widen to encompass custodians of other longitudinal data assets.

The Review also envisages that the Longitudinal Data System Custodian and Advisory Council will have a stronger focus on supporting longitudinal surveys than administrative data, at least initially. There are various administrative linkage and curation projects currently underway; the governance arrangements proposed by the Review should not interrupt this momentum. Over time, there may be a growing need for increased support and coordination of administrative data, particularly in relation to research projects focussed on administrative data and linkages. The Longitudinal Data System Custodian and Advisory Council should have the flexibility to provide this support and add value to these projects as required.

Funding

The Longitudinal Data Asset Custodians would continue to be responsible for their own funding arrangements. However, they may find it valuable to increase the coordination of future funding proposals with other asset custodians. The Longitudinal Data System Custodian could help to facilitate such coordination, such as through the development of a joint-NPP for future funding of these longitudinal data assets (as discussed above).

Priority 2. Strengthen the longitudinal data system (continued)

LONGITUDINAL DATA SYSTEM CUSTODIAN

Roles and responsibilities

The Longitudinal Data System Custodian would be responsible for facilitating, developing and communicating best practice approaches across the longitudinal data value chain to support the Longitudinal Data Asset Custodians. It would not be gatekeeper for the Longitudinal Data Asset Custodians, nor would it overtake or conflict with their existing governance arrangements. Rather, it would be a support and resource to the Longitudinal Data Asset Custodians through a 'community-of-practice' approach to increasing communication, consistency and cooperation across the longitudinal data system.

The Longitudinal Data System Custodian's responsibilities should be articulated in a Terms of Reference document to be developed and agreed to by all Longitudinal Data Asset Custodians during the implementation stage of the Review. These responsibilities could include the following:

- Facilitating the development of standards for the creation and collection of longitudinal survey data, including consistent documentation and metadata standards.
- Facilitating the development of protocols for processing and preparing longitudinal data in accordance with privacy and confidentiality requirements.
- Facilitating improvements to longitudinal data storage and access arrangements, such as by developing a 'front door' platform for accessing longitudinal data.
- Assisting with brokering data sharing and linkage arrangements between Longitudinal Data Asset Custodians.
- Coordinating the strategic plans for data assets prepared by the Longitudinal Data Asset Custodians, including by working with the Longitudinal Data Asset Custodians to optimise their data assets.

- Maintaining a catalogue of longitudinal data assets, and tracking the use and impact of longitudinal data.
- Engaging with researchers, academia and other organisations to strengthen partnerships and innovation.
- Promoting the use of longitudinal data, including by facilitating the development of innovative products that demonstrate the value of longitudinal data.
- Providing secretariat support to the Longitudinal Data System Advisory Council and preparing an annual report for the Secretaries Data Group.

Membership

The Steering Committee considers that the NCLD is best placed to initially perform the role of the Longitudinal Data System Custodian, which should continue to be housed in DSS. This will provide a smooth and quick path towards the establishment of the recommended governance arrangements, and obviate the need to go through the process of establishing a new statutory entity. The success of these arrangements can be examined in the proposed three-year review of the longitudinal data system by the Longitudinal Data System Advisory Council.

Funding

The NCLD will require additional funding to enable it to fulfil the system custodian responsibilities set out above. It is important that this funding is sufficient for the NCLD to realise the full opportunities of Australia's longitudinal data system. These funding arrangements should be formulated in a joint NPP by the Longitudinal Data Asset Custodians in the implementation stage of the Review.

Priority 2. Strengthen the longitudinal data system (continued)

LONGITUDINAL DATA SYSTEM ADVISORY COUNCIL

Roles and responsibilities

A Longitudinal Data System Advisory Council would provide oversight of the Longitudinal Data System Custodian and advice to Government on Australia's longitudinal data needs. This would include the following proposed responsibilities:

- Undertaking system planning, including advising the Government on longitudinal data investment and collection priorities. This would include ensuring that the longitudinal data system properly aligns with the Government's broader public data agenda.
- Reviewing and coordinating the strategic plans for data assets prepared by the Longitudinal Data Asset Custodians.
- Reviewing and endorsing longitudinal data standards and protocols.
- Providing strategic planning on the 'social licence' for longitudinal data in collaboration with a community working group.
- Seeking input on the longitudinal data system from other parties as required, such as data-users and international counterparts.
- Undertaking a longitudinal data system review in three years, and every five years thereafter.

In fulfilling these responsibilities, the Longitudinal Data System Advisory Council should draw upon New Zealand's approach to creating a 'data future partnership' to strengthen its data ecosystem.

Membership

The Steering Committee envisages that the Longitudinal Data System Advisory Council be constituted by a broad range of key actors in the longitudinal data system. This might include the Core National Longitudinal Data Asset Custodians, Department of the Prime Minister and Cabinet (PM&C), the Australian Statistician,

the Privacy Commissioner, state and territory government representatives, research sector representatives, and private sector representatives. It may be necessary to devise a rotating membership system to contain the total size of the Council. The details of these arrangements should be decided in the implementation stage of the Review and through an agreed Terms of Reference.

Funding

The Review expects that the Longitudinal Data System Advisory Council will require significantly less funding than the Longitudinal Data System Custodian (particularly if the System Custodian provides it with secretariat support). These funding arrangements should also be formulated through the development of a joint NPP in the implementation stage of the Review.

NEW ZEALAND DATA FUTURE PARTNERSHIP

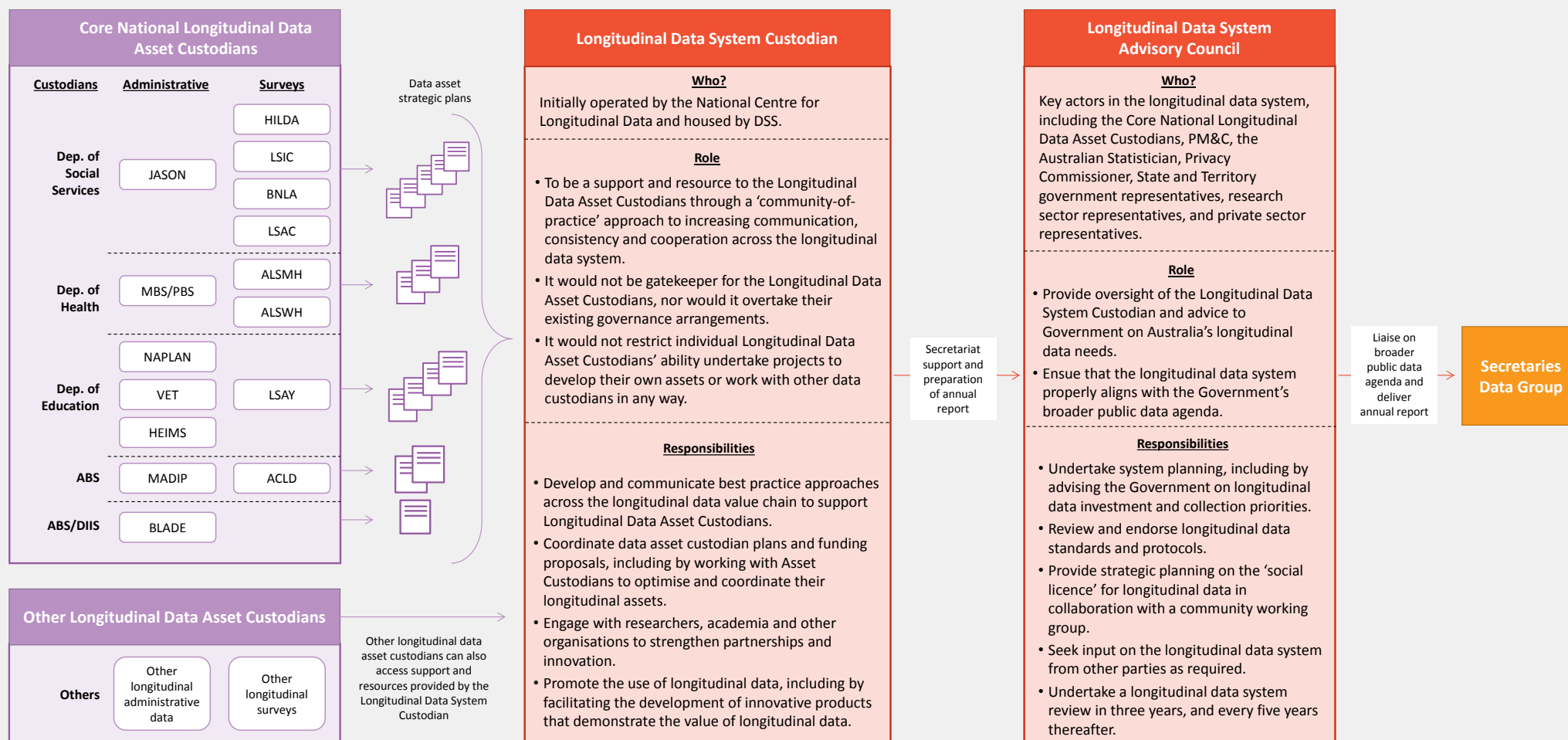
The New Zealand Data Future Partnership (NZDFP) is a cross-sector group of stakeholders who provide a collective voice on data issues, and engage with public, private and NFP sectors as well as the public to create a data-use ecosystem based on four key principles – value, inclusion, trust and control.

The NZDFP has three work streams: 1) catalyst projects to demonstrate the value of data use, and help create ethical and practical data sharing frameworks; 2) diagnose and fix ongoing and emerging issues in the data use-ecosystem; and 3) continuing to facilitate a conversation with New Zealanders about the potential value of data use.

An example of a catalyst project is Data Commons, which seeks to build whole of NZ and sector specific data sharing ecosystems that are managed peer-to-peer by key actors. In the first instance, this project is addressing the lack of platforms to safely share data within and across sectors.

Further information about the NZDFP is provided in Appendix F.

RECOMMENDED FEDERATED GOVERNANCE ARRANGEMENTS FOR AUSTRALIA'S LONGITUDINAL DATA SYSTEM



Priority 2. Strengthen the longitudinal data system (continued)

Alignment of the proposed governance model with the broader public data agenda

This proposed federated governance structure should be a part of the broader public data agenda. As shown in the figure overleaf, the longitudinal data system would operate as one component in the public data system. This will help to ensure that the longitudinal data system is properly coordinated with other data-related initiatives as part of the implementation of the Public Sector Data Management Report and the application of the Public Data Policy Statement.

This will mean that the Longitudinal Data System Custodian and Advisory Council will contribute to, but not control various elements of the public data agenda that are relevant to the longitudinal data system. For example, the development of data skills and capability is currently a strong focus of the public data agenda currently being overseen by Deputy Secretaries Data Group. This includes the development of a Data Skills and Capability Framework which will encompass data training partnerships, an APS data literacy programme and data fellowships. The longitudinal data system should not duplicate these efforts, but rather work to ensure that they encompass the specific data skills required for longitudinal data analysis.

Similarly, the longitudinal data system should support the Government's efforts to improve data access and innovation, and better coordinate data integration and sharing initiatives. These issues will also be informed by the current Productivity Commission Inquiry into Data Availability and Use.

The critical role of state and territories, the research sector and other organisations

State and territory agencies fund, collect, manage and use a significant portion of Australia's longitudinal data. It is therefore important that they are a part of the longitudinal data system as representatives on the proposed Longitudinal Data System Advisory Council.

The Review notes the potential for the longitudinal data system to develop through a national body such as the Council of Australian Governments (COAG). While there are benefits of such a forum, there is a risk that progress would stall while attempts to broker overarching arrangements were made.

Instead, there should be the capacity for certain longitudinal data issues to be progressed through COAG on a case-by-case basis. For example, in 2014 COAG agreed to publish school attendance data across jurisdictions in a consistent fashion through the *MySchool* website. Similar agreements could be reached for other longitudinal data priorities, such as transitions from juvenile justice to prison.

It is also critical that universities and research institutes play a central role in the development and coordination of the longitudinal data system. The Review proposes that they are represented on the Longitudinal Data System Advisory Council. The Review also encourages stakeholders to explore ways to increase partnerships and collaboration across all parts of the longitudinal data system. The multi-organisational and multi-disciplinary Cannabis Cohorts Research Consortium offers a good example (see below).

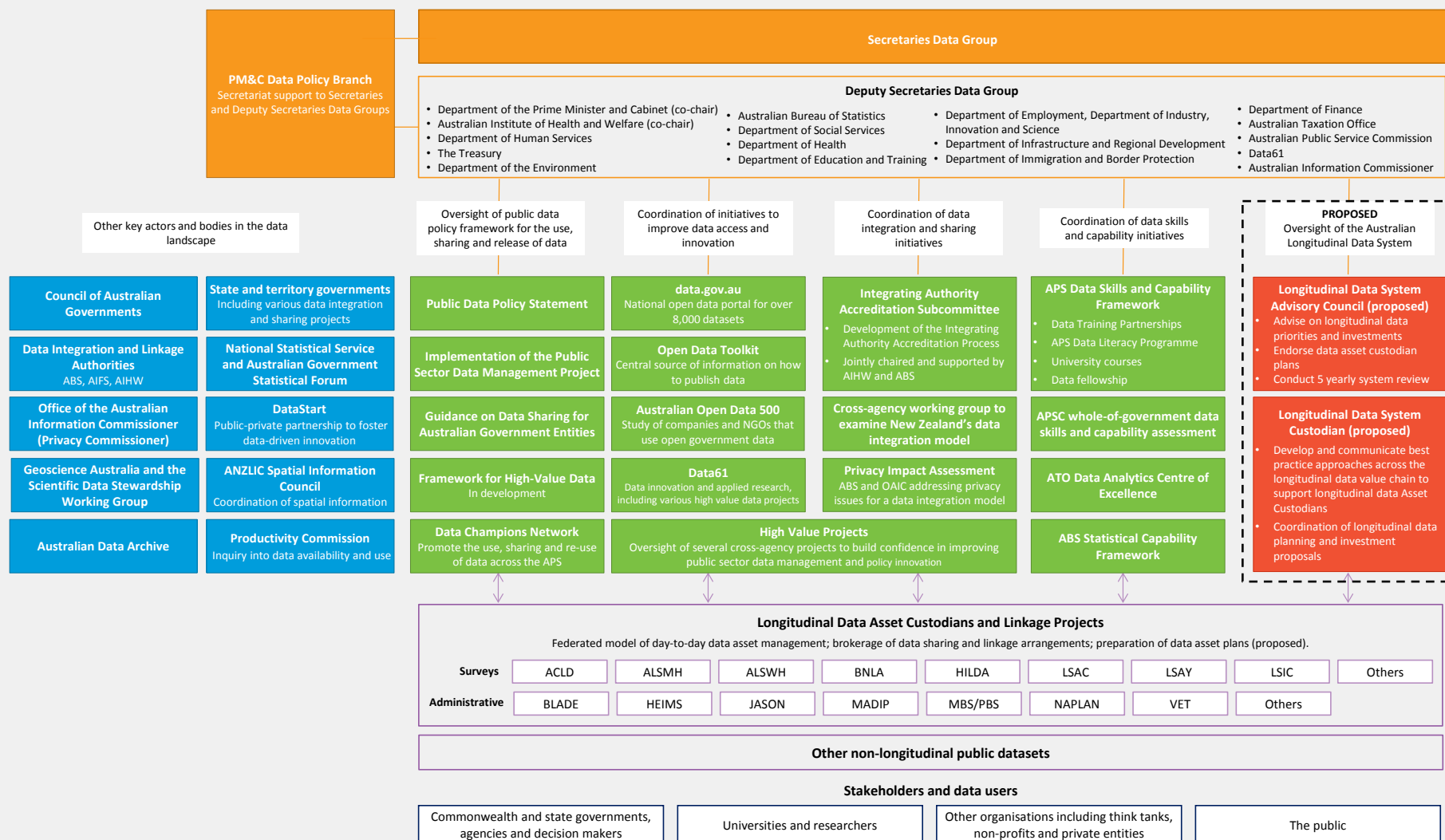
There are also a range of organisations in the non-profit and private sectors that make an important contribution to the longitudinal data system as data creators, custodians, analysts and users. These organisations should be partners in initiatives to spur innovation in longitudinal data. As such, consideration should be given to their representation on the Longitudinal Data System Advisory Council.

CANNABIS COHORTS RESEARCH CONSORTIUM

The Cannabis Cohorts Research Consortium (CCRC) is a multi-organisational and multi-disciplinary international collaboration of researchers, supported by the National, Drug & Alcohol Research Centre at UNSW Australia and the NHMRC. The CCRC stemmed from the need to better address pressing questions about the relationship between cannabis use, other drug use, life-course outcomes and mental health in children and young adults.

The CCRC brings together researchers from over 14 universities and health institutions from Australia and New Zealand. It aims to build research capacity, create the right environment to foster collaborative work practices, and build relationships between researchers involved with various longitudinal cohort studies. This collaboration has produced a large range of new research on the effects and relationships of cannabis use on children and young adults.

THE LONGITUDINAL DATA SYSTEM IN THE BROADER PUBLIC DATA MANAGEMENT LANDSCAPE



*Note: this is an indicative representation of the current public data management agenda that has been constructed by the Review. It has not been endorsed by the Deputy Secretaries Data Group or other relevant bodies. This landscape is evolving and subject to change.

Priority 2. Strengthen the longitudinal data system (continued)

Action 2.2 Create an environment that optimises access to and linkage of Australia's longitudinal data assets

The Review has identified several barriers impeding greater access to and linkage of longitudinal data. The Longitudinal Data System Custodian and Advisory Council, in partnership with other stakeholders, should explore options to overcome these barriers and strengthen the longitudinal system. The Review offers its guidance on the following opportunities to strengthen the longitudinal data system:

- develop best practice metadata standards
- improve data linkage protocols and systems
- improve data access protocols
- improve platforms for accessing and analysing longitudinal data
- contribute to reviews of privacy arrangements.

Develop best practice metadata standards

Good quality metadata is important to increase the usage of data assets. Existing approaches to metadata for longitudinal datasets vary considerably in consistency and quality. While various efforts are being made to improve these approaches for individual datasets, there is still a need to develop best practice metadata standards and protocols that will maximise the ability to align, link and analyse longitudinal datasets.

'Best practice' metadata standards would contain more detail than would be typically found in a data catalogue, including a complete and detailed description of the data and all the attributes associated with the dataset. The type of information contained in these metadata standards could include:

- description of the source of the data – and why it was collected (the source research project or department)
- the age of the data – when it was collected or how frequently it is updated or added to
- conditions or permissions given as part of collecting the data, or how the data can be used
- permissions or approvals required in order to use the data
- lists of the data elements (i.e. fields) and coding used (i.e. data dictionary).

Additional information should be documented for longitudinal surveys, including survey questions and response answers (including skip rules), coding of responses, information on how the questions have been presented (e.g. paper, CATI, online, mobile device), and links to relevant research.

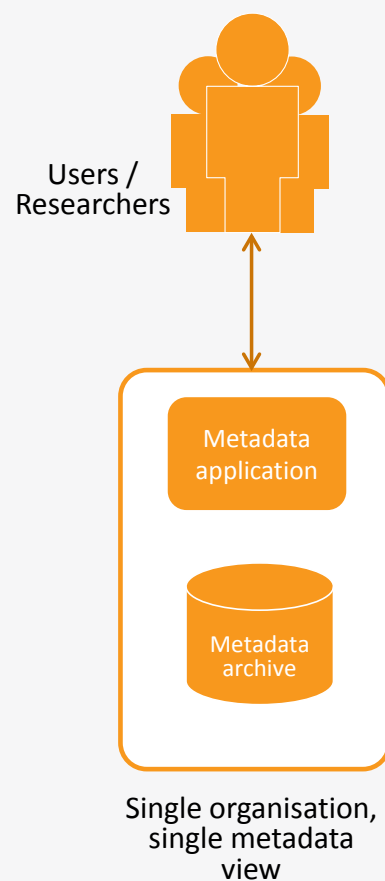
The collection of metadata could occur either through a single standalone configuration, or be configured to work in a federated design. In a standalone configuration, a single archive would draw together information provided by the data custodians. In this model the manager of the archive would be responsible for ensuring quality and comprehensiveness. In contrast, a federated approach would be dependent upon data custodians developing their own metadata documentation in accordance with agreed standards. See the diagram overleaf for a depiction of these models.

In either model, the metadata archive should be accompanied by a metadata application that enables users to undertake detailed searches of the content within Australia's longitudinal datasets. A well-functioning metadata application should enable users to identify datasets that contain identical (or near identical fields), as opposed to simply identifying datasets with similar subject matter coverage. Separately, users should be able to quickly search datasets that cover subjects of interest.

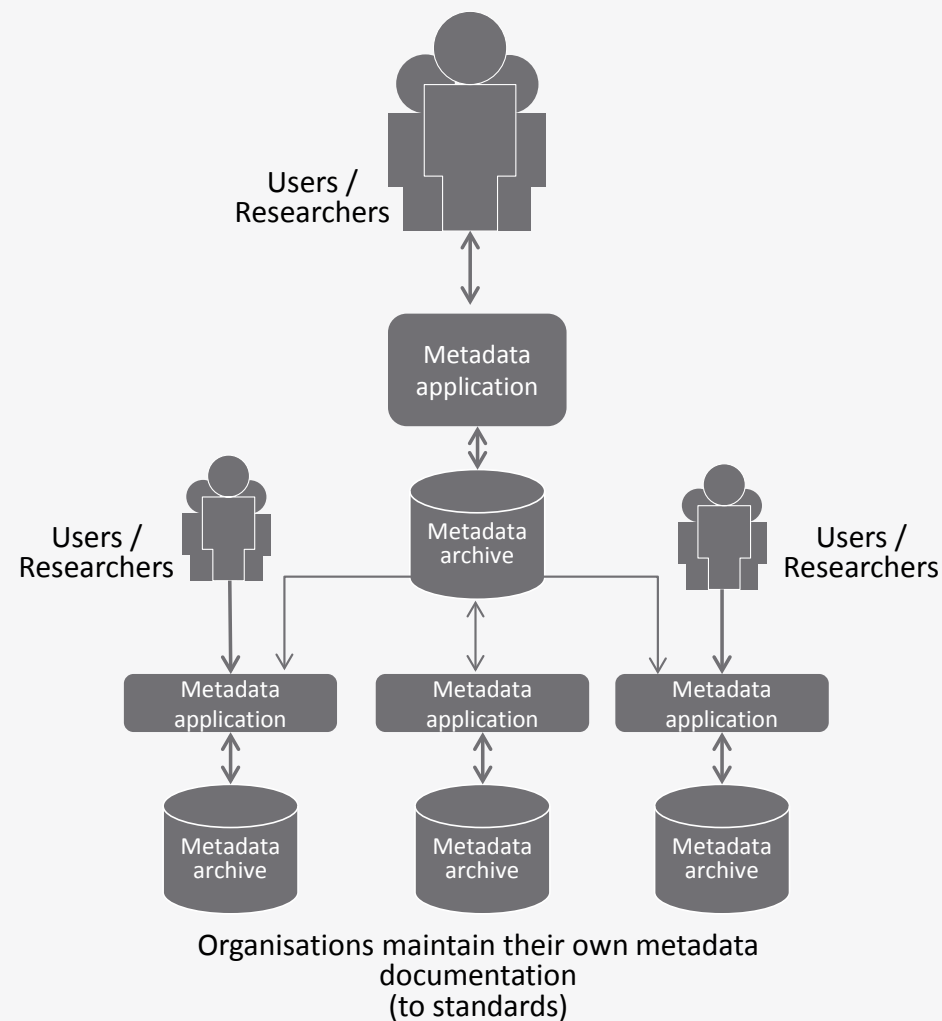
METADATA ARCHIVE OPTIONS

The diagram to the right shows the two options for a metadata archive.

Single organisation metadata model



Federated metadata model



Priority 2. Strengthen the longitudinal data system (continued)

Improve data linkage protocols and systems

Linking longitudinal data will be an increasing focus of the longitudinal data system. The complex nature of data linkage, particularly when it is necessary to generate linkage keys, means that it would be efficient to closely coordinate data linkage projects and the creation of linkage identifiers. In creating an environment that optimises linkage, thought needs to be given to the 'nimbleness' of the system so that it can support future needs such as data-on-demand, rather than a burdensome heavy infrastructure approach that is costly to government.

The Longitudinal Data System Custodian and Advisory Council should work with the Longitudinal Data Asset Custodians and other stakeholders in the public data landscape to explore these opportunities. This process should be closely aligned with the Deputy Secretaries Data Group's oversight of other data sharing and linkage initiatives, including the 'Guidance on Data Sharing Activities for Australian Government Entities' that was recently released by the Australian Government.

The Deputy Secretaries Data Group is also overseeing a Data Integration Partnership for Australia project that is exploring options for data infrastructure, building analytical capability and managing the social licence. This project is drawing on New Zealand's experience with data infrastructure, and will ensure that any new approach is fit-for-purpose for Australia's needs. The Longitudinal Data System Custodian and Advisory Council will need to be closely aligned with this project.

Improve data access protocols

The longitudinal data system should include an enabling environment that optimises the access and utilisation of longitudinal assets for government and non-government users. The Review recommends the development of a 'trusted-user' model that is consistent for all datasets. This is consistent with the Public Data Management report.

There are various ways that a 'trusted user' model could be devised. One option would be to draw on the Australian Government security clearance process, which vets individuals to obtain access to classified information and resources.

NEW ZEALAND'S APPROACH TO DATA INFRASTRUCTURE

New Zealand's Integrated Data Infrastructure (IDI) is a large research database containing de-identified microdata about people and households. Data is from a range of government agencies, Statistics NZ surveys (including the 2013 Census), and non-government organisations. The IDI holds over 166 billion facts, taking up 1.22 terabytes of space – and is continually growing. Researchers use the IDI to answer complex questions to improve outcomes for New Zealanders.

Access to the IDI is provided through a Statistics NZ 'Data Lab'. To gain access, applicants must successfully prove their project is for bona fide research purposes that are in the public interest. Researchers can only access data that is essential for their research project, and they can only access this data in secure Data Lab locations.

PROPOSED DATA ANALYTICS HUB

In its submission to the Productivity Commission's Inquiry into Data Availability and Use, DSS proposed the creation of a 'Data Analytics Hub' to link public datasets supplied by Commonwealth custodian agencies.

DSS suggested creating an "infrastructure-light, minimally resourced and nimble coordinating entity" to focus on managing appropriate data linkage arrangements and maintaining agreements between agencies, while data custodians would remain responsible for key data assurance and maintenance functions, leveraging current investments in business intelligence and physical infrastructure. The Data Analytics Hub would provide an agile 'just-in-time' service model to respond to users' needs at a whole-of-government level.

Priority 2. Strengthen the longitudinal data system (continued)

The data access tiers could range from Tier 0 for the general public to Tier 4 for government analysts accessing main unit record files (MURFs). Under this model a university student researcher (Tier 1) would have a much lower level of access to datasets than a trusted researcher based within an Australian Government research agency such as the Productivity Commission (Tier 4).

Such options should be explored in collaboration with stakeholders across the public data landscape. The Review understands that several government entities are developing options with the intent of creating a whole-of-government model. The Longitudinal Data System Custodian should play an active role in the development of these options to ensure that they are fit for purpose for longitudinal data.

Improve platforms for accessing and analysing longitudinal data

A key challenge for designing access and analysis platforms for longitudinal data is to balance the following requirements:

- Data confidentiality and privacy are preserved.
- Data is available for analysis by researchers within and outside government.
- Data analysts are able to apply longitudinal analysis techniques, ranging from tabular-based transition analysis to statistical modelling (e.g. panel regression).
- Analysis tools cater for data analysts with skills ranging from novice through to advanced users.
- Data accuracy is maintained, whereby data analysts are able to generate identical (or near identical) results using provided data, to what would be obtained using the 'original' data.
- Platforms are affordable for government and users.

The Review has been unable to identify a single secure platform that meets all the above requirements. A summary of the range of options for making data available for analysis is shown in the table overleaf.

The Longitudinal Data System Custodian and Advisory Council should work with the Longitudinal Data Asset Custodians and data users to explore the feasibility, cost and trade-offs associated with these various options. This process should draw on overseas initiatives, such as the UK Data Service, as well as domestic initiatives, such as the Remote Access Project developed by DSS and AIHW.

THE UK DATA SERVICE

The UK Data Service provides a single point of access to a wide range of secondary data including large-scale government surveys, international macrodata, business microdata, qualitative studies and census data from 1971 to 2011. The UK Data Service Nesstar catalogue provides online public access to a selection of key survey datasets in two forms: research (datasets in entirety) and teaching (dataset samplers and subsets). This includes some of the most internationally acclaimed longitudinal studies such as the British Household Panel Survey.

These datasets are complemented through the provision of tools that permit analyses, including tabulations, correlations and graphical charts. Some datasets include a 'mapping' variable that permits variables to be represented on a map of the UK. In addition, it provides multi-channel trainings for social science data users. Approved analysts can also access microdata held by the UK Data Service.

DSS AND AIHW REMOTE ACCESS PROJECT

DSS and AIHW have successfully developed a proof-of-concept project to improve data access by researchers. The project enables researchers to access data via a curated gateway. Behind the gateway is de-identified information about individuals to which queries are applied and aggregated answers extracted. Individuals' records cannot be extracted. The project will be scaled up and developed under a trusted-user model.

DATA PLATFORM OPTIONS

The table below presents the main data platform options for accessing and analysing longitudinal data.

Platform	Description						
	Analyst own environment	On-site data laboratory	Remote access computing environment	Pre-analysis confidentialisation	Post-analysis confidentialisation	Synthetic data	Custodian premises
Summary	Analyst using their own computer and server for analysis	Analyst using data custodian data laboratory	Analyst using remote access environment, with full functionality	Analyst uses a remote analysis server environment. Unit record data cannot be viewed, with data subjected to confidentialisation, and results subject to clearance	<ul style="list-style-type: none"> Analyst uses a remote analysis server environment Data is less confidentialised – may comprise MURF Analysis estimates are adjusted to reduce risk of confidentiality breach 	Analyst uses a remote analysis server environment to analyse synthetic data	Analyst uses custodian on-site facilities to undertake analysis. A more ad-hoc arrangement than an on-site data laboratory
Examples	Some surveys managed by DSS can be accessed by approved users via CD-ROM	ABS Data Laboratory (ABSDL)	<ul style="list-style-type: none"> Secure Unified Research Environment (SURE) ABS Microdata laboratory - still in development phase 	ABS RADL	ABS Tablebuilder, DataAnalyser	None in Australia US Census Bureau Survey of Income and Program Participation (SIPP) Synthetic Beta (SSB) microdata	Analyst co-location with data custodian
Confidentiality approach	Confidentialised managed through access approval process, and only making confidentialised data available	<ul style="list-style-type: none"> Access approval regime minimises attempt at breach Monitoring and vetting of all analyses Data may also be confidentialised 	<ul style="list-style-type: none"> Access approval regime minimises attempt at breach Unable to download source data, with all analysis recorded 	<ul style="list-style-type: none"> Confidentialised unit data not viewable Analysis results subject to review Access approval regime may also minimise attempt at breach 	<ul style="list-style-type: none"> Minimise ability to breach privacy Access regime may also minimise attempt at breach 	<ul style="list-style-type: none"> Minimise ability to breach privacy Access regime may also minimise attempt at breach 	All analysis is reviewed by custodian
Data location	User storage	Custodian server	Virtual server of data custodian	Virtual server of data custodian	Virtual server of data custodian	Virtual server of data custodian	Data custodian
Functionality	Software capability of user ranging from statistics package to data analysis tools (e.g. Excel or Cognos, with pre-built reports)	Applications offered by data laboratory	Dependent on applications offered by custodian	<ul style="list-style-type: none"> Dependent on applications offered by custodian RADL has limited functionality (e.g. no graphing) 	Dependent on applications offered by custodian	Dependent on applications offered by custodian. In the US Census Bureau example, service is provided for post-analysis validation using MURF	Dependent on applications offered by custodian
Estimate accuracy	Typically low - dependent on data confidentialisation	Typically high, especially if MURF used	Typically high, especially if MURF used	Medium level – dependent on extent of data confidentialisation	Typically high, especially if MURF used	Medium level – depends on analysis methods used	Typically high, especially if MURF used
Ease of use	Users own software	Dependent on applications offered - typically advanced user	Dependent on applications offered - typically advanced user	Dependent on applications offered - typically advanced user	Tablebuilder: novice; DataAnalyser: advanced	Dependent on applications offered	Dependent on custodian applications

Priority 2. Strengthen the longitudinal data system (continued)

Contribute to reviews of privacy arrangements

The ABS and the Office of the Australian Information Commissioner are conducting a privacy impact assessment for issues associated with the adoption of a data integration model. In addition, the Productivity Commission's Inquiry into Data Availability and Use will likely examine various privacy related issues.

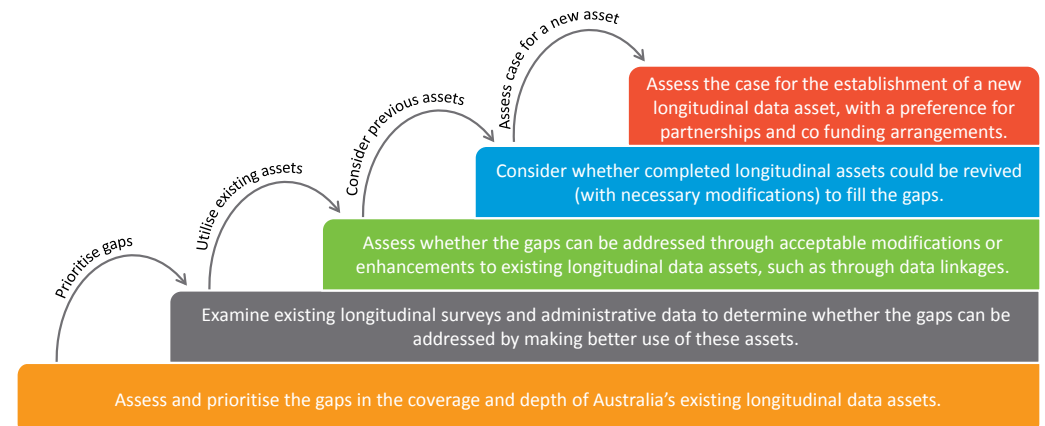
The stakeholders in the longitudinal data system should contribute to these activities by providing advice on privacy issues affecting the access and linkage of longitudinal data. One option could be to propose a broadening of the public interest exemptions in the *Privacy Act 1998*. This would be consistent with the 2008 Australian Law Reform Commission of Inquiry (ALRC Report 108), which recommended that the health and medical research exemptions extend to human research more generally, and that 'research' include the compilation or analysis of statistics.

Priority 3. Invest in the harmonisation, coverage and impact of longitudinal data

Action 3.1 Improve the efficiency, effectiveness and coverage of Australia's Core National Longitudinal Data Assets

While Australia has a good base of longitudinal data, there are significant opportunities to increase its coverage and depth across the life-course and various public policy issues. Appendix E provides an overview of the main gaps in the coverage of the Core National Longitudinal Data Assets. These gaps have been synthesised from a combination of the Review's analysis and stakeholder consultations. It should be used as an indicative starting point for further testing and analysis.

The Review recommends that the Longitudinal Data System Custodian and Longitudinal Data System Advisory Council engage in the process set out in the diagram below to provide ongoing advice to Government on future longitudinal data investments.



Priority 3. Invest in the harmonisation, coverage and impact of longitudinal data (continued)

Improve platforms for accessing and analysing longitudinal data

Consideration should also be given to opportunities to harmonise and increase the efficiency of managing longitudinal data assets to enable reinvestments in the system. This might include:

- increasing the interoperability of longitudinal data assets through more consistent use of data design and collection
- exploring the feasibility of integrating the 'back office' of longitudinal assets
- increasing the use of technology and innovation in the collection and management of longitudinal data
- using data linkages to achieve some efficiencies over the medium to long term in survey design and collection
- examining the trade-off between survey cost, sample sizes and the frequency of survey waves
- exploring options to better coordinate the promotion and use of longitudinal data.

These possibilities require detailed analysis and testing by the Longitudinal Data Asset Custodians, in collaboration with the Longitudinal Data System Custodian and Advisory Council.

Action 3.2 Promote the use of longitudinal data by demonstrating its value through insights and products that are meaningful to decision makers

One of the key system weaknesses identified by the Review is that longitudinal data is underutilised by decision makers and not well understood outside of expert circles. The Longitudinal Data System Custodian should therefore play a role in promoting the use of longitudinal data by demonstrating its value to decision-makers and other stakeholders. This might include:

- commissioning research projects specifically focussed on longitudinal data assets
- developing new and innovative products that demonstrate the value of longitudinal data
- promoting the use of longitudinal data in evaluations and reviews
- tracking the use and impact of longitudinal data in policymaking and academic research, including by keeping a catalogue of research papers that make significant use of longitudinal data.

These projects should demonstrate the 'power' of longitudinal data across the value chain – from linking data, to robust analysis, to presenting policy relevant results in compelling ways. The products should utilise a variety of mediums, ranging from reports through to interactive tools. They should have an emphasis on data visualisation techniques to present insights from longitudinal analysis in a meaningful and digestible way.

These initiatives should not be undertaken by the Longitudinal Data System Custodian in isolation. Rather, it should collaborate with other partners across the public data landscape. In particular, there should be an increased focus on engaging and partnering with the research sector to demonstrate the value of longitudinal data.

Priority 3. Invest in the harmonisation, coverage and impact of longitudinal data (continued)

There is also a need to identify and promote the range of skills and occupations required to optimise Australia's longitudinal data system. The Review considers that advanced longitudinal data analysis skills are required both within and outside government, both to validate findings from longitudinal data and to use the data to undertake ongoing and policy-relevant analysis.

The Longitudinal Data System Custodian should contribute to the data skills agenda being coordinated by the Deputy Secretaries Data Group to ensure that it includes a focus on the particular skills needed for longitudinal data analysis. This should include the diverse range of skills required across the longitudinal data value chain, from information technology and programming (for data extraction and database management), mathematics (for data linkage), and statistics and econometrics (for data analysis). Additional skills are required to translate insights from longitudinal analysis for a policy-focussed audience. The increasing use and prevalence of 'big data analytics' will also require the development of skills and tools that permit machine learning and artificial intelligence analytics.

EXAMPLES OF PROJECTS DEMONSTRATING THE VALUE OF DATA TO THE PUBLIC POLICY MAKING PROCESS

Behavioural Economics Team of the Australian Government

The Behavioural Economics Team of the Australian Government (BETA) within PM&C is a joint initiative across government to use insights from behavioural economics to help improve policy outcomes. It will undertake various demonstration projects and seek to use longitudinal data to understand real -human behaviour and place it at the centre of policy and program design.

Data Analytics Centre of Excellence

The Data Analytics Centre of Excellence was established by the ATO as a space to build analytics capability across government. The purpose of the Centre of Excellence is to enable a common capability framework for analytics, as well as an opportunity to share technical knowledge, skills and tools. The Centre of Excellence will also help build collaborative arrangements with tertiary institutions to aid the development of analytics professionals.

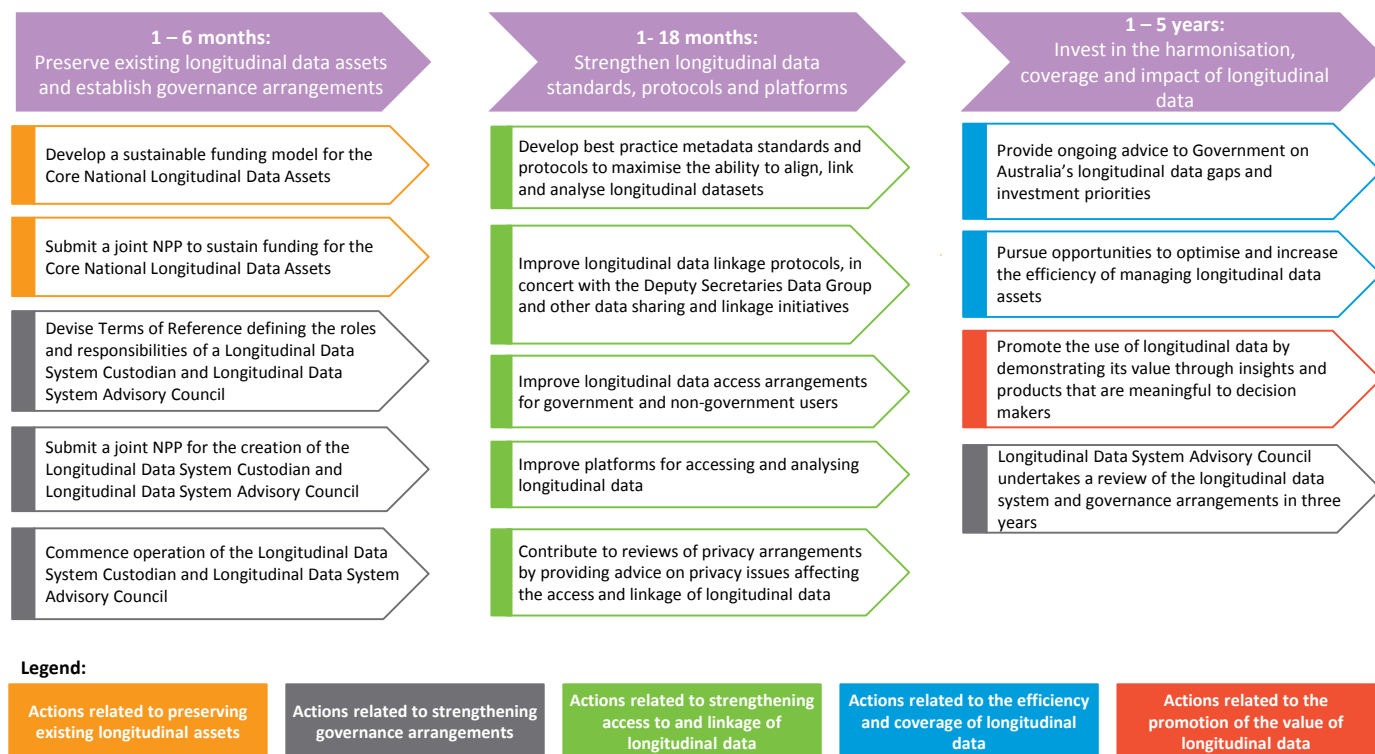
Business Longitudinal Analytical Data Environment

The Department of Industry, Innovation and Science is currently undertaking a project to improve the BLADE (described above in Section 3), including its scope and accessibility and thereby its capacity to provide data for government and private industry to better understand Australian firms and industry. This project will expand the range of data accessed via the BLADE, and the use of technology solutions to enhance access and security (including access by researchers).

12. Roadmap for implementation

The Review suggests that an implementation working group be established to advance the recommended priorities in this report, with a particular focus on preserving funding for the Core National Longitudinal Data Assets and establishing the governance arrangements for the longitudinal data system. This working group could be composed of the Steering Committee for this Review, or just the Longitudinal Data Asset Custodians. As discussed above, the NCLD is well placed to play the role of the initial Longitudinal Data System Custodian during this establishment phase and lead the coordination of the implementation working group.

The implementation working group should devise a Terms of Reference for the governance arrangements and agree on matters such as reporting lines, interrelationships with other governance arrangements, processes for the resolution of disputes and the sustainability of governance structures. It should also consider developing joint NPPs to create these governance arrangements and sustain existing longitudinal data assets. A roadmap for the implementation of the Review's recommended priorities is provided to the right.



APPENDICES

Appendix A Terms of Reference

Terms of Reference

The Australian Government has commissioned a review of longitudinal data architecture as part of the “Investment Approach to Welfare” measure in the 2015-16 Budget. The key objective of the review is to inform Australia’s future longitudinal data needs.

Background

From July 2015, the Government will develop and begin implementing an ‘investment approach’ to Australia’s social security system. This is consistent with the recommendations of the McClure Review of Australia’s Welfare System and builds on lessons learned from the New Zealand investment model and the United Kingdom’s ‘try, test and learn’ approach.

An optimal investment approach will require high quality and integrated administrative data and longitudinal survey data. Longitudinal data will be critical to conducting robust valuations and evidence-based interventions.

The McClure Review states that longitudinal data is vital to a better understanding of life course transitions for policy development and service design (p. 222). Integrated longitudinal administrative data can provide detailed information about government transfers to and from individuals and families but lacks detailed information about individual behaviours and circumstances that can be obtained through survey data. Together, longitudinal administrative and survey data provide a rich source of data to inform policy.

Current longitudinal data collections are often developed in a segmented and ad-hoc way. This ignores the strong interrelationships between factors across policy domains. Furthermore, the analytical potential of existing data has not yet been fully realised and longitudinal administrative data is underutilised.

This review will examine the architecture of Australia’s current longitudinal data assets in order to identify strengths and weaknesses and thus inform future data collection.

Scope of the Review

A cross-departmental committee of SES officers will lead the review. This committee will be supported by the National Centre for Longitudinal Data (NCLD), drawing on external, technical expertise as required. The review will:

- develop a strategic plan for the Australian Government’s longitudinal data architecture to support evidence-based policy development
- promote a coordinated approach to longitudinal data developments, informed by the needs of government, academia and the private sector
- consider arrangements for supporting longitudinal data integrity, development, analysis and dissemination in Australia
- consider the current and potential use of longitudinal administrative data, the complementary role of longitudinal surveys, and how best to balance investments in these data assets
- propose a framework prioritising investment by the Commonwealth in new and existing longitudinal survey and administrative data assets.

As part of this work, the review will examine the current state of longitudinal surveys and administrative datasets in Australia, and advise on how best to ensure that our longitudinal data assets are of a high quality, and can best meet the needs of data users across government, academia and the private sector. This consideration will include:

- overview of the value of different types of longitudinal data to inform policy
- investigation of gaps and overlaps in the current suite of longitudinal datasets and the potential for these data to inform long term policies to improve the lifetime wellbeing of people and families across Australia
- an inventory of Commonwealth, state and institutional longitudinal surveys, and relevant administrative datasets, to help inform the gap analysis
- assessment of the complementary strengths of longitudinal survey data, in the context of Australia’s cross-sectional, time series and administrative data assets

-
- 
- the most valuable potential data integration projects, both between different longitudinal administrative datasets and between administrative and longitudinal survey data
 - assessment of arrangements needed to:
 - i. support a more coordinated approach to the development of Australia's longitudinal data investments
 - ii. build relationships and collaborations between policy agencies and academic centres with longitudinal data expertise, both in Australia and internationally
 - iii. develop and implement longitudinal data quality standards, including, where appropriate, data harmonisation to facilitate cross-survey, cross-national and cross-cohort comparisons
 - iv. develop data collection standards, including advising on best practise for managing respondent and community engagement, consent, privacy and ethics
 - v. promote and facilitate greater use of administrative data linkage
 - vi. develop best practices for data security, storage, access and release
 - vii. promote the development and maintenance of longitudinal data analysis and item design capabilities
 - viii. promote the accessibility and useability of longitudinal data and increase the number of data users into the future
 - ix. support the translation and dissemination of longitudinal data analysis to inform policy and ensure that policy thinking, in turn, informs the ongoing development of, and investment in, longitudinal data.
 - structure, coherence and overlaps of the suite of current longitudinal studies managed by the NCLD and assessment of how well these studies align with best practise standards and processes outlined above.

Process

The NCLD will be undertaking an appropriate consultation process, including a series of workshops and engagement activities with stakeholders from government, academia and the private sector. Scoping, planning, assessing and implementing tasks are planned to be undertaken in 2015-16. A preliminary report on findings will be disseminated towards the middle of 2016, and the review will be finalised in 2016.

Appendix B Steering Committee and stakeholders consultations

Steering committee	Stakeholders consulted	
Australian Government agencies Mr Sean Innis (Chair) – Special Adviser, Productivity Commission (previously Group Manager, Policy Office, Department of Social Services) Dr Tim Reddel – Group Manager, Policy Office, Department of Social Services Mr David Dennis – Branch Manager, Policy Evidence Branch, Department of Social Services Mr Adam Rowland – Executive Manager, National Centre for Longitudinal Data Ms Alana Foster – First Assistant Secretary, Research Data and Evaluation Division, Department of Health Ms Mary McDonald – Group Manager, Evidence and Assessment, Department of Education and Training Ms Jo Wood – Group Manager, Economic Strategy, Department of Employment Ms Michelle Wilson – General Manager, Strategic Information, Health and Information Group, Department of Human Services Ms Bridget Brill – A/General Manager, Strategic Information, Health and Information Group, Department of Human Services (replaced Ms Michelle Wilson) Ms Elizabeth Hefren-Webb – First Assistant Secretary, Schools, Information and Evaluation Division, Department of the Prime Minister and Cabinet Dr Paul Jelfs – General Manager, Population & Social Statistics Division, Australian Bureau of Statistics Mr Robert Ewing – Principal Advisor, Tax Analysis Division, The Treasury Dr Jenny Gordon – Principal Adviser Research, Productivity Commission Tertiary Institutions and Research Institutes Professor Deborah Cobb-Clark – University of Sydney Professor Stephen Zubrick – University of Western Australia, Faculty of Education Professor Mick Dodson AM – Director of the National Centre for Indigenous Studies, Australian National University Associate Professor Daryl Higgins – Deputy Director, Australian Institute of Family Studies Private sector Ms Kate Inglis-Clark – Design Review Evaluate	Commonwealth Government Departments Department of Social Services National Centre for Longitudinal Data Department of Health Department of Education and Training Department of Employment Department of Human Services Department of the Prime Minister and Cabinet Attorney General's Department Department of Industry, Innovation & Science Department of Immigration and Border Force The Treasury Other Commonwealth Government entities Australian Bureau of Statistics Productivity Commission Australian Taxation Office Australian Research Council Australian Institute of Health & Welfare Office of the Australian Information Commissioner Australian Institute of Family Studies Data 61 (Commonwealth Scientific and Industrial Organisation) Office of the Chief Scientist State government Departments and Agencies New South Wales Behavioural Insights Unit New South Wales Department of Finance New South Wales Department of Education Western Australia Department of Education Western Australian Department of Health New South Wales Data Analytics Centre SA-NT Datalink	Tertiary Institutions DVCR Australian National University (ANU) DVCR University of Western Australia ANU (National Centre for Epidemiology and Population Health) ANU Centre for Aboriginal Economic Policy Research ANU School of Demography ANU Tax and Transfer Policy Institute ARC Centre of Excellence for Children and Families over the Life Course University of Melbourne - Melbourne Institute of Applied Economic and Social Research University of Tasmania Research Institutes and Think Tanks Telethon Kids Institute Murdoch Childrens Research Institute Centre of Excellence in Population Ageing Research Australian Statistics Advisory Council Australian Research Alliance for Children and Youth Sax Institute National Institute of Labour Studies Other Roy Morgan Research Australian Council of Social Service Centre for Social Impact International Canada Statistics Canada Canadian Institute for Health Information Manitoba Centre for Health Policy New Zealand New Zealand Data Alliance New Zealand Superu New Zealand Privacy Commissioner New Zealand Data Future Partnership Loyalty New Zealand – LAB360 New Zealand Treasury – Social Investment Unit Statistics New Zealand

Appendix C Examples of the value of longitudinal data

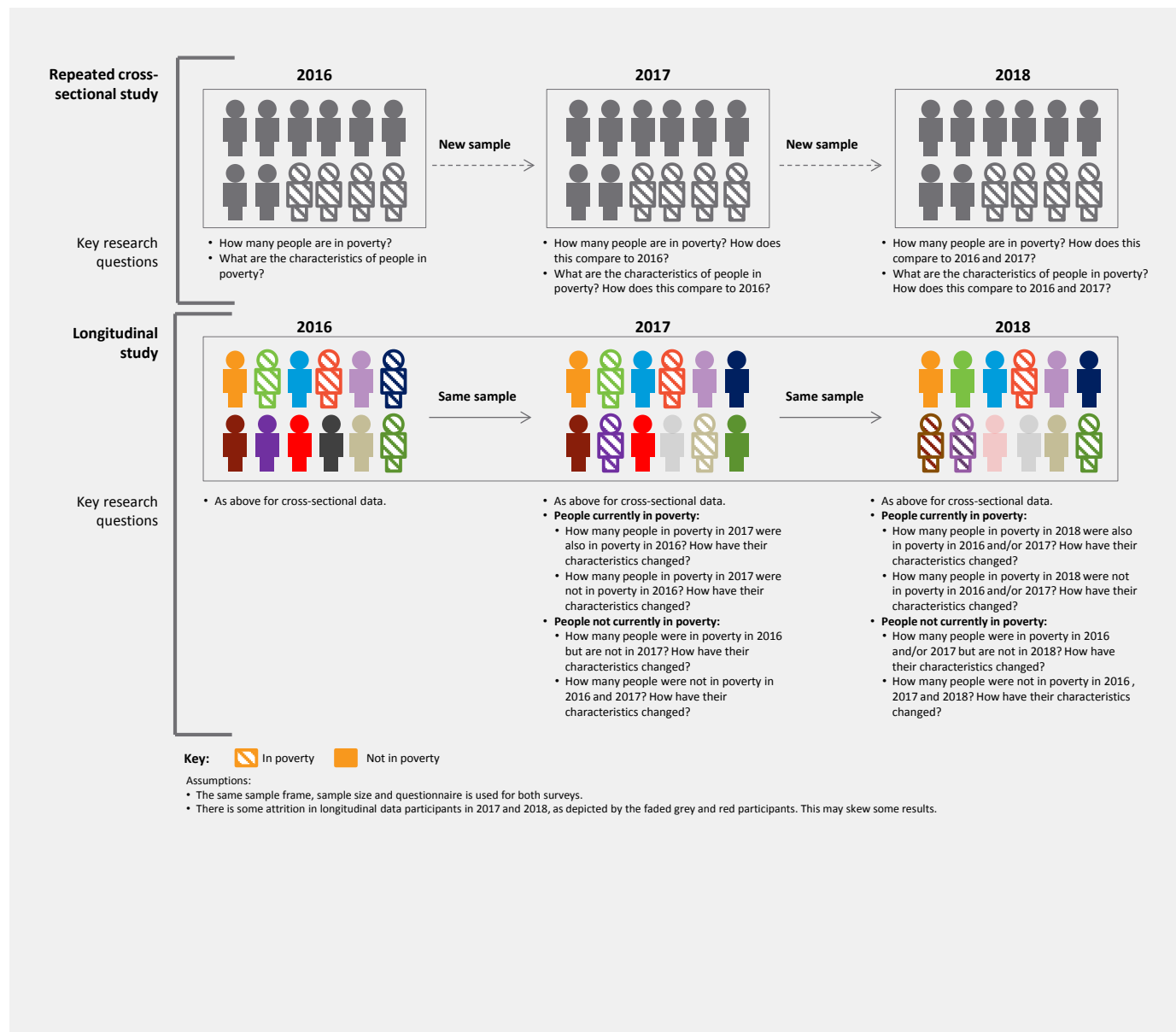
STYLISTED EXAMPLE

What can longitudinal studies tell us about poverty?

The figure to the right presents a stylised comparison of the sorts of insights about poverty that we can derive from repeated cross-sectional and longitudinal studies.

Both surveys tell us how many people are in poverty, what their characteristics are, and how this has changed over time (level and trend insights). The cross-sectional study has the advantage of being able to refresh its sample every time it runs its survey, which means that the results may be more representative of the changing population. Additionally, the cross-sectional study does not have to deal with participants dropping-out and skewing the data.

But only the longitudinal survey tracks the same individual over time. This can be used to tell us about the transitions people make into and out of poverty over time. By comparing their characteristics and circumstances to other people in the sample, we can start to gain a better picture of the factors that cause and reduce poverty.



APPLIED EXAMPLE

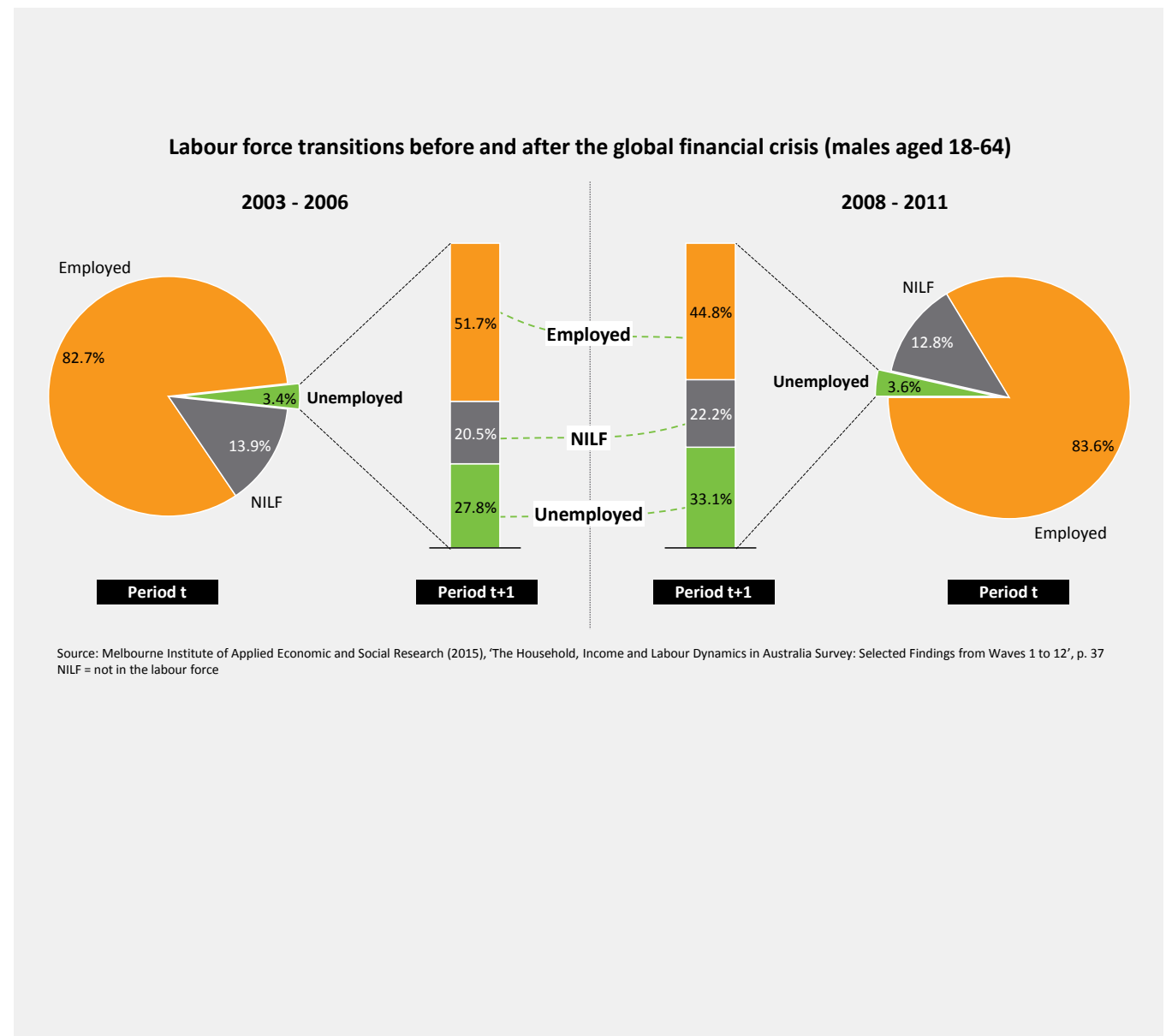
What can longitudinal studies tell us about unemployment?

The figure to the right provides an applied example of the value of longitudinal data through an analysis of labour force transitions using the HILDA survey. The example examines one-year labour force status transitions before and after the Global Financial Crisis (GFC) for males aged 18 – 64.

A cross-sectional analysis shows that in both time periods – 2003 to 2006, and 2008 to 2011 – the proportion unemployed men was nearly identical at 3.4% and 3.6% respectively. This suggests that the GFC had little impact on unemployment.

However, a longitudinal analysis using HILDA data can provide insight into what was really happening in the labour market over these periods. This analysis shows that men who became unemployed during the GFC were less likely to transition out of unemployment than those who experienced unemployment before the GFC.

In the pre-GFC period (2003 to 2006), 27.8% of men who were unemployed in a previous year remained unemployed a year later. In the GFC period (2008 to 2011) 33.1% of men who were unemployed a previous year remained unemployed. A greater proportion of men who became unemployed during the GFC also dropped out of the labour force compared to the earlier period.






International and domestic examples of the impact of longitudinal data on public policy

There are many celebrated examples of longitudinal studies changing the way we understand society and its problems. These insights have often translated into new policy applications that have greatly improved well-being, both in Australia and abroad. Some prominent examples are highlighted below.

INTERNATIONAL

 Longitudinal study	 Example insights	 Policy application
British Birth Cohort Studies (UK, 1946 – ongoing) A birth cohort study of British people born in 1946, 1958, 1970, 1991 and 2000.	<p>Disadvantaged women in 1946 were 70% more likely to have stillbirths than less disadvantaged women.</p> <p>Smoking during pregnancy is associated with reduced birth weight and a higher risk of infant mortality.</p>	<p>The launch of the National Health Service in 1948 included free medical care associated with pregnancy and birth.</p> <p>Change in advice to women on the risks of smoking during pregnancy.</p>
British Doctor's Study (UK, 1951 – 2001) A longitudinal cohort study of practicing UK doctors to understand the link between smoking and various diseases.	<p>One of the earliest and most influential studies to present convincing statistical proof of the causal relationship between smoking and increased mortality rates. The 50-year follow up of the same physicians confirmed the suspected relation of smoking to 12 of 13 types of cancer.</p>	<p>Formed the foundation of global public health debates over smoking in the 1960s and 1970s, leading to increased tobacco regulation and taxes.</p>
HighScope Perry Preschool Study (USA, 1962 – 2005) A longitudinal study of 123 disadvantaged children in Michigan that received high-quality preschool programs. The study followed the participants until age 40. Additional data was gathered from the subjects' school, social services, and arrest records.	<p>Participants at age 40 who attended the preschool program had higher earnings, were more likely to hold a job, had committed fewer crimes, and were more likely to have graduated from high school than the participants who did not attend the preschool program.</p>	<p>The study formed the basis of programs that started in the 1980s to provide comprehensive early childhood education to low-income US children and their families. In 2013 the US Congress introduced a bill (Strong Start for Children Act) to further expand access to high quality, full-day preschool for 4-year-olds from low to moderate income families.</p>
Multicenter AIDs Cohort Study (MACs) (USA, 1985 – ongoing) An ongoing prospective study of 7,000 men with and without HIV in Baltimore, Chicago, Pittsburgh and Los Angeles.	<p>The study helped to determine that AIDS is a viral illness and provided evidence on how it is transmitted.</p>	<p>MACs has informed global public health policy on HIV education, prevention and treatment.</p>

DOMESTIC

 Longitudinal study	 Example insights	 Policy application
Australian Longitudinal Study on Women's Health (ALSWH) (1996 – ongoing) A longitudinal survey of Australian women in various age cohorts which assesses their physical and mental health, psychosocial aspects of health (such as socio-demographic and lifestyle factors) and their use of health services.	<p>The study has identified specific prevalence and predictors of weight gain, sedentary behaviours and health, and new domains of physical activity for Australian women.</p>	<p>Informed numerous federal and state government policies across a wide range of issues. For example, the Australian Government's Physical Activity and Sedentary Guidelines for Adults draws on ALSWH to make recommendations about minutes of walking per week and the amount of exercise required to avoid weight gain for older women.</p>
Longitudinal Surveys of Australian Youth (LSAY) (1995 – ongoing) A cohort study that follows young people as they move from school into further study, work and other destinations. Participants enter the study when they turn 15 years and are contacted once a year for 10 years.	<p>LSAY is the pre-eminent source of quantitative information for describing the Australian youth transition process and determinants, and also sheds light the characteristics of 'at-risk youths'.</p>	<p>Informed several Government reviews including the Career Transitions and Partnerships Strategic Review, OECD Jobs for Youth Thematic Review, Youth Compact and National Partnership on Youth Attainment and Transition, and COAG Reform Council reporting on youth transitions.</p>
Wittenoom Cohort Studies (1975 – ongoing) A retrospective study of men employed at Wittenoom between 1943 and 1966. In 2002, the Wittenoom Cohort Study expanded to women who had lived and worked in the town.	<p>Conclusively established the dangers of even relatively slight exposure to blue asbestos, and continues to inform scientific research into the causes, treatments and potentially vaccine to mesothelioma.</p>	<p>Contributed to the 2003 Australia-wide ban on the manufacture and use of all types of asbestos.</p>

Appendix D Catalogued longitudinal datasets

Core National Longitudinal Data Assets	Other longitudinal datasets	
	Ongoing surveys	Completed surveys
Surveys <ol style="list-style-type: none"> 1. Australian Census Longitudinal 5% Dataset (2006 and 2011) 2. Household, Income and Labour Dynamics in Australia study (HILDA) 3. Longitudinal Study of Australian Children (LSAC) 4. Longitudinal Surveys of Australian Youth (LSAY) 5. Australian Longitudinal Study on Women's Health (ALSWH) 6. The Longitudinal Study on Male Health (LSMH) 7. Longitudinal Study of Indigenous Children (LSIC) 8. Building a New Life in Australia - the Longitudinal Study of Humanitarian Migrants (BNLA) Administrative <ol style="list-style-type: none"> 9. Multi-Agency Data Integration Project (MADIP) 10. Business Longitudinal Analytical Data Environment (BLADE) 11. Longitudinal Dataset for the Investment Approach (JASON) 12. Medicare Benefits Scheme/Pharmaceutical Benefits Scheme 10% Data Set 13. National Assessment Program – Literacy and Numeracy (NAPLAN) 14. National VET Data Collection 15. Higher Education Information Management System (HEIMS) 	<ol style="list-style-type: none"> 1. 2000 Stories Victorian Adolescent Health Cohort Study (VAHCS) (VIC) 2. 45 and Up study (NSW) 3. Aboriginal Birth Cohort Study (ABC) (NT) 4. Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (AIBL) (VIC; WA) 5. Australian Longitudinal Study of Adults with Autism Spectrum Conditions (ALSAA) 6. Australian Multiple Sclerosis Longitudinal Study (AMSLS / MS Life Study) 7. Australian Science Enrolment Project (NSW) 8. Australian Temperament Project (ATP) (VIC) 9. Beyond 18: The Longitudinal Study on Leaving Care (VIC) 10. The Childhood to Adolescence Transition Study (CATS) (VIC) 11. Crossroads Undiagnosed Disease Study (VIC) 12. Environments for Healthy Living (EFHL) (NSW; QLD) 13. Florey Adelaide Male Ageing Study (FAMAS) (SA) 14. Improving Management by Participatory Research in Oncology: the Victorian Experiment (IMPROVE) (VIC) 15. International Youth Development Study (IYDS) (VIC) 16. Life Patterns 17. Longitudinal Study of Behavioural and Emotional Disturbance in People with Intellectual Disability (Helping Young People Grow) (VIC) 18. Mater-University of Queensland Study of Pregnancy (MUSP) (QLD) 19. Medicine in Australia: Balancing Employment and Life (MABEL) 20. Older Australian Twins Study (OATS) (NSW; QLD; VIC) 21. Pathways of Care: Longitudinal Study on Children and Young People in Out-of-Home care in New South Wales (POCLS) (NSW) 22. Personality and Total Health Through Life (PATH) (ACT) 23. SA Dental Longitudinal Study (SA) 24. Social Futures and Life Pathways of Young People in Queensland (Our Lives) (QLD) 25. Study of Australian Students with Autism (LASA) 26. Sydney Centenarian Study (SCS) (NSW) 27. Tasmanian Longitudinal Health Study (TAHS) (TAS) 28. The Australian Longitudinal Epilepsy Study (ALES) (VIC) 29. Bettering the Evaluation and Care of Health (BEACH) 30. The Busselton Healthy Ageing Study (BHAS) (WA) 31. The Concord Health and Ageing in Men Project (CHAMP) (NSW) 32. The Koori Growing Old Well Study (KGOWS) (NSW) 33. The Melbourne Longitudinal Studies on Health Ageing Program (MELSHA) (VIC) 34. The Study of Environment on Aboriginal Resilience and Child Health (SEARCH) (NSW) 35. The Western Australian Pregnancy Cohort Study (Raine Study) (WA) 	<ol style="list-style-type: none"> 36. A Longitudinal Study of Bone Loss in Men (SA) 37. Adult Migrant English Program Longitudinal Study (AMEP) 38. Australia at Work 39. Australian Child to Adult Development study (ACAD) 40. Australian Diabetes, Obesity and Lifestyle (AUSDIAB) 41. Australian Longitudinal Study of Ageing (ALSA) (SA) 42. Australian Longitudinal Study of Health and Relationships (ALSHR) 43. Blue Mountains Eye Study (BMES) (NSW) 44. Canberra Longitudinal Study of Ageing (CLSA) (ACT; NSW) 45. Childhood Determinants of Adult Health (CDAH) 46. Diabetes and Related Disorders in Urban Indigenous People in the Darwin region (The DRUID Study) (NT) 47. Diagnosis, Management and Outcomes of Depression in Primary Care (Diamond study) (VIC) 48. Family Pathways: The Longitudinal Study of Separated Families (LSSF) 49. Health Of Young Victorians Study (HOYVS) (VIC) 50. Journeys Home 51. Longitudinal Survey of Immigrants to Australia (LSIA) 52. Millennium Mums Survey 53. Negotiating the Life Course (NLC) 54. Paid Parental Leave Evaluation 55. Participation in cervical screening by Indigenous women in the Northern Territory (NT) 56. Restorative Justice and the Life Course: Victims and Offenders in Longitudinal Perspective 57. Stronger Families in Australia (SFIA) 58. Sydney Memory and Ageing Study (SMAS) (NSW) 59. Sydney Older Persons Study (SOPS) (NSW) 60. The Health in Men Study (HIMS) (WA) 61. The Hunter Community Study (HCS) (NSW) 62. The Wittenoon Cohort Studies (WA) 63. Victorian Gambling Study (VIC)

Appendix E Key gaps in the Core National Longitudinal Data Assets

Policy theme	Sub-policy theme	Current longitudinal data assets	Potential longitudinal data gap
Education & Employment	Early Childhood Development, Schools	<ul style="list-style-type: none"> Longitudinal Study of Australian Children Medicare Benefits Scheme/Pharmaceutical Benefits Scheme 10% dataset Personality and Total Health Through Life Australian Temperament Project Aboriginal Birth Cohort Study Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing Australian Longitudinal Study of Adults with Autism Spectrum Conditions The Childhood to Adolescence Transition Study Longitudinal Study of Behavioural and Emotional Disturbance in People with Intellectual Disability Study of Australian Students with Autism 	<ul style="list-style-type: none"> Birth cohorts: LSAC may require a new cohort that begins at the prenatal stage, as well as funding to continue following its current cohorts through youth and adulthood. Child health: Limited information on the population-wide impact of interventions on long-term health. Integration: Longitudinal tracking of individuals across the health care system, encompassing diagnosis, health care utilisation, associated expenditure within and across health care sectors (e.g. primary care, acute care), including private health insurance.
	Schools, VET, Higher Education	<ul style="list-style-type: none"> Longitudinal Study of Australian Children Longitudinal Surveys of Australian Youth Higher Education Information Management System 2000 Stories Victorian Adolescent Health Cohort Study Australian Science Enrolment Project Beyond 18: The longitudinal study on leaving care National VET Data Collection 	<ul style="list-style-type: none"> Tracking school students: Limited ability to track students across the stages of education (pre-school, primary and secondary school, vocational education, and higher education). Some tracking of school-age children is possible within jurisdictions, but this is impeded if they move between jurisdictions and/or school sectors. LSAC provides some data but does not answer all relevant research questions. Key data of interest includes participation, attendance, skills (e.g. NAPLAN) and attainment. Examples of high value metrics not currently able to be measured include: <ul style="list-style-type: none"> participation rate of children of pre-school age in an early childhood education program (reflecting the fact that many children may attend both childcare and pre-school) national (or state/territory) year 12 completion rates (only apparent rates are reported). Linkages: Limited linkage of educational data with: <ul style="list-style-type: none"> other policy priorities (e.g. post-education employment outcomes, family resilience etc.), including repayment of HELP debt. data relating to out-of-school influencers (e.g. maternal educational status, aspirations, peer group).
	Employment	<ul style="list-style-type: none"> Australian Census Longitudinal 5% Dataset (2006 and 2011) Household, Income and Labour Dynamics in Australia study Longitudinal Surveys of Australian Youth Business Longitudinal Analytical Data Environment Life Patterns Medicine in Australia: Balancing Employment and Life 	<ul style="list-style-type: none"> Unit-level integration: Lack of unit-level administrative data on personal income and wealth (assets and liabilities), and employment.

Policy theme	Sub-policy theme	Current longitudinal data assets	Potential longitudinal data gap
Health	Ageing	<ul style="list-style-type: none"> Australian Longitudinal Study on Women's Health 45 and Up study Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing Florey Adelaide Male Ageing Study Older Australian Twins Study Sydney Centenarian Study The Busselton Healthy Ageing Study The Concord Health and Ageing in Men Project The Melbourne Longitudinal Studies on Health Ageing Program 	<ul style="list-style-type: none"> Retirement well-being: Limited information on what affects wellbeing in retirement and how this changes over the retirement life-course. Retirement decisions: Limited information on the factors influencing decision-making of soon-to-be or existing retirees. Mobility: Limited information on mobility and its implications. Bequests: There is insufficient information on bequests.
	Reducing the Gap in Indigenous Disadvantage	<ul style="list-style-type: none"> Longitudinal Study of Indigenous Children Aboriginal Birth Cohort Study The Koori Growing Old Well Study The Study of Environment on Aboriginal Resilience and Child Health 	<ul style="list-style-type: none"> Life course coverage: No comprehensive 'life course' coverage of Indigenous Australians due to a lack of a dedicated adult Indigenous longitudinal survey. There is a question as to whether longitudinal data on Indigenous Australians can be sourced from sufficient samples in other surveys, or whether a stand-alone survey is required. Representativeness: Longitudinal Study of Indigenous Children is not fully representative and may consequently be underutilised.
Society & Community	Interpersonal violence	<ul style="list-style-type: none"> Pathways of Care Australian Longitudinal Study on Women's Health 	<ul style="list-style-type: none"> Domestic violence: Poor longitudinal survey and administrative data coverage of domestic violence, encompassing both victims and perpetrators. There is also scope for linkage with social security data, and other collections.
	Housing and Homelessness	<ul style="list-style-type: none"> Beyond 18: The longitudinal study on leaving care 	<ul style="list-style-type: none"> Housing: Inadequate longitudinal survey and administrative data coverage of key issues related to housing tenure, home purchasing and ownership, housing stress, and associated taxation and investment issues (e.g. negative gearing). Journeys Home, a national survey on housing and living challenges has completed all six waves.
	Family resilience	<ul style="list-style-type: none"> Longitudinal Study of Indigenous Children Household, Income and Labour Dynamics in Australia study Longitudinal Study of Australian Children Building a New Life in Australia The Childhood to Adolescence Transition Study Pathways of care Social Futures and Life Pathways of Young People in QLD 	<ul style="list-style-type: none"> Family formation and deformation: Limited information on the factors influencing family formation and deformation. Linkages: Limited integration to permit tracking of individuals and families related to child protection services and crime, being in receipt of social security.
	Immigration	<ul style="list-style-type: none"> Building a New Life in Australia Australian Census Longitudinal 5% Dataset (2006 and 2011) 	<ul style="list-style-type: none"> Immigrants: There exists only one longitudinal survey - Building a New Life in Australia – Longitudinal Study of Humanitarian Migrants - but this does not cover non-humanitarian migrants.
	Crime and Justice	<ul style="list-style-type: none"> Pathways of Care 	<ul style="list-style-type: none"> Access: Longitudinal administrative data on interactions with criminal justice system (police, courts, corrections) is largely unavailable. Only some jurisdictions have the capacity to extract such data.

Appendix F International case studies

What is the jurisdiction's longitudinal data context?	What are the leading examples of this jurisdiction's longitudinal data approach?	Why is this important?
CANADA		
Canada is considered one of the earliest adopters of statistically-informed policy making. Since the 1990s, it has undertaken numerous initiatives to better link up data, primarily for the purposes of informing policy development and evaluation (and less focussed on providing wider public access to its data holdings)	<p>In 1994, Statistics Canada created the Longitudinal Administrative Databank (LAD). The LAD is a longitudinal file designed as a research tool on income and demographics. It comprises a 20% sample of the annual T1 Family File (T1FF) and the Longitudinal Immigration Data Base. The T1FF comprises summary income tax and welfare benefits data.</p> <p>LAD data are extracted from administrative files and derived from other Statistics Canada (ABS equivalent in Canada) surveys and other sources. It contains many annual demographic variables about the individuals represented and their census family in that year, covering variables related to cultural background, family, economic wellbeing, spending and taxation. For immigrants landed between 1980 and 2012, the file also contains certain key characteristics observed at landing.</p> <p>Variables have been harmonised where possible and individuals can be linked year to year starting with 1982 data. The file is augmented annually with new data, permitting deeply detailed analysis of Canadians over their lifetime.</p>	<p>The LAD is impressive in its breadth of coverage (and consequential design) and the extent of historical information it can provide on tax filers.</p> <p>The longitudinal nature of the LAD permits custom-tailored research into dynamic phenomena, as well as representative cross-sectional patterns. The LAD microdata is made available for analysis in secure Research Data Centres (RDCs), and is mainly used by government departments to evaluate programs and support policy recommendations. Academics, private consultants and Statistics Canada researchers also use the data for analyses of socio-economic conditions.</p>
NEW ZEALAND		
For the last two decades, New Zealand has increasingly sought to better link and integrate its national data assets. New Zealand has been exploring and pursuing a centralised approach to creating and managing longitudinal data, whereby investment and focus on longitudinal data is driven by the country's overarching 'longitudinal data needs'	<p>The New Zealand Data Future Partnership (NZDFP), a cross-sector group of influential individuals who provide a collective voice on data issues, is mandated by the NZ Government to engage with public, private and NFP sectors as well as New Zealanders to create a data-use ecosystem based on four key principles – value, inclusion, trust and control. Specifically, the NZDFP seeks to address issues with the trusted data use including limited cross-sectoral collaboration, tenuous social licence, complex data practices and relationships and the underutilization of data in creating social and economic value.</p> <p>The NZDFP has three work streams. 1) catalyst projects to demonstrate the value of data use, and help create ethical and practical data sharing frameworks, 2) diagnose and fix ongoing and emerging issues in the data use-ecosystem and 3) continuing to facilitate a conversation with New Zealanders about the potential value of data use. An example of a catalyst project is Data Commons, which seeks to build whole of NZ and sector specific data sharing ecosystems that are managed peer-to-peer by key actors.</p>	<p>NZDFP is an important example for the Australian context for two reasons.</p> <p>First, it seeks to harness the opportunities for trusted data sharing across private, public and not-for-profit groups with central consideration for its citizens.</p> <p>Related to this, the NZDFP is founded under the principle that it must build and maintain social licence to collect, access and use data. It seeks to do this by continually engaging with stakeholders to understand perceptions, and its catalyst projects.</p>
UNITED KINGDOM		
The UK has sought to facilitate high quality social and economic research and education through promoting sharing and to encourage the best-practice reuse of its extensive longitudinal administrative and survey data holdings.	<p>The UK Data Service provides a single point of access to a wide range of secondary data including large-scale government surveys, international macrodata, business microdata, qualitative studies and census data from 1971 to 2011.</p> <p>The UK Data Service Nesstar catalogue specifically provides online public access to a selection of key survey datasets in two forms: research (datasets in entirety) and teaching (dataset samplers and subsets). This includes some of the most internationally acclaimed longitudinal studies such as the British Household Panel Survey.</p> <p>These datasets are complemented through the provision of tools that permit analysis, including tabulations, correlations and graphical charts. Some datasets include a 'mapping' variable that permits variables to be represented on a map of the UK. In addition, it provides multi-channel trainings for social science data users. Approved analysts can also access microdata held by the UK Data Service.</p>	<p>The UK Data Service is a leading example of providing open access and tools to assist longitudinal analysis at all levels of public, private and NFP sectors.</p> <p>Everyday, its data holdings are accessed in the interest of:</p> <ul style="list-style-type: none"> • advancing research (e.g. analysing the causes of the gender pay gap) • informing public policy (e.g. the impact on alcohol pricing policies on crime and health) • adding relevance and interest to college and university coursework through the use of real-life (teaching) datasets (e.g. applied economic theory)

Appendix G Glossary of terms

Term	Definition
Administrative data	Information collected for delivering public administration – for example to conduct registration, transaction and record-keeping activities.
Cohort study	A form of panel study. A cohort study is focussed on a group of individuals who share a fixed characteristic or have shared an event during a particular time span – for example all children born in a particular year.
Core National Longitudinal Data Assets	Datasets that the Review has identified as being ‘core’ to Australia’s national public longitudinal data holdings. These datasets meet a set of criteria related to coverage, representativeness, quality and accessibility (amongst others).
Cross-sectional data	Data collected by observing many subjects (such as individuals or households) at the same point in time (or over the short period of data collection).
Data analysis and research	The fourth step in the longitudinal data value chain. Involves analysis for the needs of government and external researchers.
Data analysts	Entities that are funding to undertake primary analysis of longitudinal datasets and/or access data and undertake analysis for their own purposes.
Data assets	Survey or administrative datasets.
Data collection	The second step in the longitudinal data value chain. Involves collecting either survey data or administrative data.
Data creators	Entities responsible for primary data collection.
Data custodians	Entities responsible for the management, storage and security of a dataset.
Data custody	The third step in the longitudinal data value chain. Involves data extraction, processing, linkage (optional), storage and security and access.
Data end users	Entities and people that use longitudinal data insights to conduct research and inform the public policy making process.
Data funders	Entities that provide financial or other support for a longitudinal dataset and studies.
Data linkers	Entities that integrate two or more datasets.
Data matching	The process of creating linked data.
Data planning	The first step in the longitudinal data value chain. Involves longitudinal data prioritisation, design, purchasing and set up of evaluation.
Data sharing	The transfer of data between organisations.
De-identified data	Data relating to a specific individual where the identifiers have been removed to prevent identification of that individual. Otherwise known as anonymised data.
Designed data	<i>See survey data.</i>
Enduring policy themes	A set of policy themes that the Review has identified as being enduring and high priority for current and future governments. These include health, education and employment, and society and community.

Term	Definition
High-value data	Data which would have a higher economic value if made available as open data due to its authoritativeness, timeliness, accuracy or other traits.
Life-course relationship ready,	A framework through which to analyse the coverage of longitudinal data across over seven 'life stages': pre-natal/child ready, school ready, life ready, work ready, retirement ready, and end of life ready.
Linked data	Data created from matching and integration of two or more datasets. This may occur through either an explicit match on unique identifiers, or through a combination of information that gives a high confidence match between the datasets.
Longitudinal data	Data collected from the same individuals, households or entities at several points in time – the data is designed so that differences within and between individuals can be tracked over time.
Longitudinal data architecture	The infrastructure and enablers required for the collection and creation, capture and analysis of longitudinal data. This includes data extraction, rules, processing, documentation, linkage, storage, and the range of analysis using longitudinal data. Also included is the organisational structures needed to support these activities.
National Centre for Longitudinal Data	A business unit within the Department of Social Services, responsible for promoting a longitudinal evidence base that informs policies and practices to improve the lifetime wellbeing of people and families in Australia.
Non-sensitive data	Data that is anonymised and does not identify an individual or breach privacy or security requirements.
Open data	Open data refers to making datasets available so that others can use them without restriction on use or redistribution in its licensing conditions.
Organic data	<i>See administrative data.</i>
Panel study	A type of study that tracks a sample of individuals or households and is typically focussed on the dynamics of economic and subjective well-being, employment and families.
Personal data	As defined by the Privacy Act 1988, data relating to a specific individual where the individual is identified or identifiable in the hands of a recipient of the data.
Public sector data	Data collected or generated by the public service for policy development and public administration. Also known as Public Sector Information.
Re-identified data	Data where the identity of an individual has been ascertained from a de-identified dataset – usually through comparison with other datasets.
Repeated cross-sectional data	Where the same data items are measured using separate samples of the population taken at two or more points in time – the data does not normally measure the same individuals at each point. For this reason, this type of data can measure aggregate changes in a population but cannot measure changes within individuals over time (as with longitudinal data, described above).
Survey data	Data sourced from a survey. The survey may be undertaken by a face-to-face or telephone interview, or be provided by the respondent using an internet-based or paper questionnaire. A survey is typically undertaken using a sample of the population.

Term	Definition
Time series data	A type of repeated cross-sectional data. A time series is a collection of observations of well-defined data items obtained through repeated measurements over time.
Trusted user	A trusted user is someone who is authorised with the appropriate security clearances and confidentiality agreements to access more sensitive public data,, such as unit-record administrative data, for purposes such as research or policy development. Examples include public servants and members of research institutions.
Unit-record data	Data that is at the most granular level – for example unit-record patient data would contain personal data about the individual patient.
Longitudinal data value chain	The five step process involved with creating longitudinal data. These steps are data planning, data collection, data custody, data analysis and research, and data use.
Waves	A period of data collection, for example, in the Longitudinal Study of Australian Children (LSAC), wave 1 occurred in 2004 when families were interviewed for the first time; Wave 2 occurred in 2006 when the same families were interviewed for the second time.