
1 Using administrative data to achieve better policy outcomes

All levels of government hold data for administrative purposes. These data sets cover large parts of the population, offering a largely untapped opportunity to evaluate policies and programs and develop more effective and efficient ones. Unlike many other countries, Australia makes relatively little use of its public data resources even though the initial costs of making data available would be low relative to the future flow of benefits. International experience shows that confidentiality can be protected, and domestically, researchers have used de-identified Western Australian data for over 30 years without any breaches of privacy.

Academics, researchers, data custodian agencies, consumers and some Ministers are eager to harness the evidentiary power of administrative data, but this enthusiasm generally is not matched by policy departments. Despite tentative steps, overall progress has been inadequate. Leadership and commitment is required to promote the evidence-based policies needed to meet Australia's economic and social objectives within budget constraints that will become more acute given the demographic outlook.

Effective policy making rests on evidence

Systematic evidence-based analysis is an essential element of all good policy. It is particularly important for social services with such a major share of budget outlays. For 2013-14, Australian Government spending is expected to be \$398 billion with social security and welfare, \$138 billion (35 per cent); health, \$65 billion (16 per cent) and education, \$30 billion (7 per cent) (Australian Government 2013). Australia-wide, expenditure on health alone was around \$130 billion in 2010-11 of which the Australian and state and territory governments funded 69 per cent (AIHW 2012). Significantly, the costs of health and aged care are expected to rise sharply with Australia's ageing population and advances in medical treatments.

The Commission has previously addressed the need to strengthen evidence-based policy development (PC 2010b). It postulated that community expectations of what governments can do about policy problems often run ahead of reality or are

influenced unduly by sectional interests. In Australia, this can be compounded by failure to draw on information that would elucidate understanding of problems and proposed solutions. The Commission identified several contributing factors:

- a diminution, over many years, of specialist public sector research bureaux
- in-house evaluations, to the extent they are done, being conducted by policy departments that are constrained in the frankness of their (public) evaluations
- relatively little experience of public agencies sharing data with academics and other external specialists (PC 2010b).

A rich vein of information is held by governments in the form of ‘administrative data’ collected for regulatory requirements (e.g. vehicle registrations and taxation declarations), program administration (e.g. Centrelink and Medicare payments, school, university and vocational enrolments and completions, and hospital admissions) or as a byproduct of transactions (e.g. fines and fees) (ABS 2011, p. vi).

The Commission concluded that access to de-identified data for government users, academics and other researchers should be pursued as a priority (PC 2010b). But its recent work is testimony that gaining access to administrative data remains difficult. In *Caring for Older Australians*, the Commission noted that:

... given that the Government already collects and maintains detailed data sets relating to aged care, the provision of better public access to this data is likely to generate sizable net benefits... the default presumption should be that data be transparent and automatically released in a timely manner. (PC 2011a, pp. 462-3)

Similarly, in *Disability Care and Support* it considered:

Data are a key aspect of the evidence base of a good insurance scheme (and badly lacking in the current disability system) ... (PC 2011b, p. 564)

A Commission staff paper on *Deep and Persistent Disadvantage* found:

Administrative data has the potential to provide new knowledge to inform researchers and policy makers about ... disadvantage. (McLachlan et al., 2013, p. 2).

Administrative data (and data matching) is commonly used to detect undeclared income by welfare recipients (McLucas 2013) or over-claiming by service providers. While these initiatives reduce waste of scarce resources and reinforce public confidence, the savings from improved program integrity are likely to pale in comparison to the costs that can arise if the underlying programs themselves are poorly designed and therefore less effective. Used for comprehensive policy analysis, data matching could identify programs that do not work and for whom and where enhancements could be made to programs that do. Making these data available would enable independent verification of official evaluations, as well as providing insights of relevance to governments at low cost.

Administrative data are sources of evidence

Australia is well positioned to take advantage of its administrative data resources:

- all Australian governments hold extensive longitudinal administrative databases containing high quality information about large populations
- increases in computing power, data storage and data capture and matching technologies mean that analysis of very large databases is increasingly feasible
- advances in analytical techniques allow investigation in ways that can isolate policy impacts from other influences (Leigh 2010, Smith and Sweetman 2010)
- the Objects of the *Freedom of Information Act 1982* declare that information held by the government is a national resource to be managed for public purposes.¹

Yet, Australia's experience remains one of untapped potential. In 2008, Australian Government Treasury officials reported:

Having clearly defined administrative data is all very well, but it's next to useless if these data are not shared with those best able to build the evidence base. Our universities and research institutes are teeming with people wanting to draw lessons from agencies' statistics... Researchers are often forced to fumble around like the drunk that searches for his keys under a street light — not because his keys are likely to be there, but because it's the only spot where he can see. (Gruen and Goldbloom 2008)

Five years on, Professor Gregory lamented the:

... long standing government institutional failures to make the necessary data available to allow Australians to understand how their IS [income support] system interacts with the labour market Independent researchers have not been given sufficient access to administrative longitudinal IS data from Centrelink, any access to administrative data on job finding services and implementation of job seeker activation from DEEWR and any access to unit record ABS time series data ... (Gregory 2013, p. 6)

Australian researchers have often had to look elsewhere to obtain the data necessary to investigate public policy matters. The Australian Government has funded research organisations and the Australian Bureau of Statistics (ABS) to develop longitudinal databases — in areas as diverse as children, migrants, youth, ageing, and families — and make confidentialised unit record data available to registered users. The most significant broad longitudinal survey in recent years is the Household, Income and Labour Dynamics (HILDA) survey at the Melbourne Institute. Gregory (2013, p. 6) considered HILDA to be the 'the most important data innovation of the last decade'. Similarly, the Commission has noted that, by giving

¹ The Office of the Australian Information Commissioner's Principles on Open Public Sector Information states that open access to information should be the default position (OAIC 2012).

researchers access to longitudinal data, HILDA has stimulated substantial important policy relevant research (PC 2010b). The same can be said for the longitudinal surveys of Australian Children and of Australian Youth.

Are longitudinal surveys a substitute for administrative data?

While necessity has driven Australian researchers to develop different sources of evidence, surveys and administrative data are not necessarily substitutes — each has strengths and weaknesses. An advantage of surveys is the control researchers have over content at the specification stage. This is conducive to survey questions being built around soundly constructed theories and methods.

On the other hand, surveys are less likely to include sufficient numbers of particular groups, such as the most disadvantaged (e.g. homeless people or those with substance abuse problems) who are by nature difficult to contact and who may not give consent to participate. For example, when the Commission surveyed people receiving counselling for gambling problems, the majority indicated that prior to seeking counselling they would not have answered a population survey about gambling (PC 1999). And, even if surveys could initially capture a reasonable cohort of such households, this group is more likely to drop out, so apparent trends can be confounded by attrition. Apart from selection bias, survey responses may be influenced by behavioural changes that arise from the act of participation itself. Conducting surveys and seeking participants' consent can be very expensive compared to analysing existing data.

Administrative data encompass longitudinal structures that enable analysis of outcomes over time; large samples, sometimes full populations, that allow rarer events or smaller groups to be studied; and high quality information that does not suffer from rising non-response rates, attrition and under reporting. All of this adds to greater statistical power for robust policy analysis. Of course, 'raw' administrative data have characteristics that may need to be addressed if the information is to be used for policy analysis (table 1.1).

Data linkage can consolidate administrative data with information held elsewhere, such as surveys. Administrative data can indicate *what* happened to whom in terms of pathways and outcomes benchmarked against policy variations. Surveys can elicit more targeted information on *why* people behaved as they did. A further benefit of data matching would be to enable surveys to omit sensitive questions, such as income levels, substance abuse or other factors that typically get a low response. This would reduce costs and respondent burden.

Table 1.1 Advantages and disadvantages of using administrative data

<i>Advantages</i>	<i>Disadvantages</i>
<ul style="list-style-type: none"> • Collected for operational purposes, so no additional collection costs, but will incur extraction and cleaning costs 	<ul style="list-style-type: none"> • Information collected is restricted to data for administrative purposes and limited to users of services and administrative decisions
<ul style="list-style-type: none"> • Collection not additionally intrusive to target population 	<ul style="list-style-type: none"> • Lack of researcher control over content
<ul style="list-style-type: none"> • Regularly, sometimes continuously, updated 	<ul style="list-style-type: none"> • Proxy indicators sometimes have to be used
<ul style="list-style-type: none"> • Can provide historical information and allow consistent time-series to be built up 	<ul style="list-style-type: none"> • May lack contextual/background information
<ul style="list-style-type: none"> • Collected in a consistent manner, if part of a national system 	<ul style="list-style-type: none"> • Changes to administrative procedures can change definitions and make comparisons over time problematic
<ul style="list-style-type: none"> • Subject to rigorous quality checks 	<ul style="list-style-type: none"> • Missing or erroneous data. Possible incentive to fabricate responses to access benefits.
<ul style="list-style-type: none"> • Near full coverage of population of interest 	<ul style="list-style-type: none"> • Quality issues with variables may be less important (e.g. address details not updated)
<ul style="list-style-type: none"> • Reliable at the small area level 	<ul style="list-style-type: none"> • Metadata — lacking or of poor quality
<ul style="list-style-type: none"> • Counterfactuals / controls can be selected post hoc 	<ul style="list-style-type: none"> • Data protection issues
<ul style="list-style-type: none"> • Captures those who may not respond to surveys 	<ul style="list-style-type: none"> • Access by researchers dependent on support of data providers.
<ul style="list-style-type: none"> • Potential for data sets to be linked to produce powerful research resources 	<ul style="list-style-type: none"> • Underdeveloped theory and methods

Source: Smith et al. 2004.

What could be done with greater access to data?

Administrative datasets could be instrumental in gaining insights into whether government programs:

- meet their stated objectives — do they work or are other influences at play?
- operate as intended — do recipients respond to (dis)incentives and are there unanticipated (good or bad) effects on recipients or the community?
- are delivered effectively — are there queuing or discouragement effects?
- deliver services in the right places — are services located near people in need?

Such information is fundamental to deeper questions about whether the policy mix is coherent or whether other policy initiatives work to hinder desired outcomes. There may be interactions between disparate factors that impinge on outcomes which can only be detected using large data sets. Administrative data could also be used proactively to instigate debate on matters of public importance that would otherwise fail to gain traction without corroborating evidence. These benefits are increasingly recognised. The Australian Government’s ‘big data’ Issues Paper identified that processing and integrating administrative data has the:

... potential to transform service design and delivery so that personalised and streamlined services, that accurately and specifically meet individual's needs, can be delivered to them in a timely manner. (Commonwealth of Australia 2013, p. 4)

In a similar vein, this year the (former) Minister for Human Services championed the cause of better use of administrative data resources:

... if you start from the premise that you are serious about evidence-based policies you realise you can actually develop them by using the data you've already got. We know where people live, we know when they've worked and how they've responded to major shocks. We know what illnesses they have suffered, and how they were treated. We can follow a family's journey right down the generations. I want to open up that information to researchers ... For example I would like to know what type of medical admissions take place ahead of applications for child support. If we knew that, we would know where to best direct resources before they were needed. (Carr 2013)

De-identified administrative data collections could be made available to researchers, to encourage examination of policies. Robust evidence of policy efficacy need not be the sole province of sophisticated techniques like randomised control trials (RCTs) — the so-called 'gold standard' of evidence, used extensively in the United States for policy evaluation. Because RCTs can be costly, difficult to design well, and can raise ethical issues about risks for the 'treatment' or 'control' groups, they have rarely been used in Australia. If administrative data were disclosed, analysis using alternative methodologies could shed light on policy performance.

We could better understand disadvantage

Access to administrative data would provide much needed insights of the paths into, through, and out of, disadvantage. McLachlan et al. observed that:

Government agencies, at all three levels of government, hold very large administrative data sets which may assist in unlocking a deeper understanding of the factors influencing disadvantage, the government programs that are accessed by those experiencing disadvantage, and how those programs assist (or hinder) those who are the most vulnerable. (McLachlan et al. 2013, p. 196)

Using administrative data, researchers could derive evidence on people's lifecycle use of income support (Newstart, disability or other benefit), the duration(s) of use and their parents' benefit history. By linking data on other factors — such as location, educational attainment, mental health, hospitalisations and incarceration — it would be possible to analyse the pathways for individuals and families with characteristics that make them vulnerable to persistent or intergenerational disadvantage. Administrative data could identify events such as job loss, incapacity and family breakdown that contribute to individuals' transition to social exclusion. Absent this information, policy must rely on partial analyses and intuition.

We could connect more dots in health

Australia has population-based data on Medicare services, dispensing of subsidised pharmaceuticals, emergency department presentations, hospital admissions, aged care and deaths. Linked, these data have huge potential for policy-relevant research. Professor Stanley has claimed that access to real-time prescription and birth data could have detected the connection between the morning sickness drug thalidomide and thousands of birth defects much earlier.

The whole reason we set up birth defects registries across Australia was to pick up the next thalidomide.. But until now we haven't been able to link those registries to the Pharmaceutical Benefits Scheme. It's insane. (Stanley 2012)

Stanley's research also established that a maternal diet rich in folic acid can prevent spina bifida in babies. Integrating administrative data was pivotal for this work.

One study that linked MBS, PBS and Western Australian hospital morbidity data examined the scope to achieve better integrated services (DHAC 2000). The study recommended using unique patient records to automate data collection for health care monitoring. There appear to have been few subsequent studies that have been able to access and link MBS and PBS data for research.

Greater linking of health and non-health data sets could save lives and deliver more efficient and better targeted services. In 2009, the National Health and Hospitals Reform Commission recommended that:

To better understand people's use of health services and health outcomes across different care settings, we recommend that public and private hospital episode data should be collected nationally and linked to MBS and PBS data using a patient's Medicare card number. (NHHRC 2009, p. 21)

However, current privacy guidelines mean that MBS and PBS information may be disclosed for medical research, but not statistical research.² Medical research can result in more effective treatments, whereas 'statistical' research may result in programs that reduce the likelihood of conditions developing, and more efficient targeting of resources where treatments are necessary. Protecting confidentiality is warranted but the current approach is too cautious and complex with the restrictions creating unnecessary downsides and delays for evidence-based policy formulation.

2 Legislation outlining how and when Medicare Benefits scheme (MBS) and Pharmaceutical Benefits scheme (PBS) data can be linked is contained in the *National Health Act 1953* (s. 135AA and 135AB). It prohibits the storage of MBS and PBS data in the same data base and any linkage unless the linkage is specified in privacy guidelines. The Privacy Guidelines for the MBS and PBS were last issued in 2008. MBS and PBS information may be disclosed for medical research, but not statistical research, either with consent from the individuals involved or in accordance with guidelines issued by the National Health and Medical Research Council.

We could analyse the interactions between welfare and work

The pathways between welfare and work are complex. There are poverty traps arising from the effective marginal tax rates confronting those deciding to transition from welfare to work. There are also interactions with minimum wages, educational attainment, skills, location and labour mobility. There is also debate about how the level of income support affects incentives to seek work.

On the latter question, Professor Gregory sought to evaluate Australia's 'make work pay' approach by asking whether increasing the relative poverty of income support recipients leads them to increase their employment sufficiently to offset the poverty-creating element of the policy. Gregory concluded that independent research has not been able to address such questions, citing the inability to access administrative data and observing that:

... good researchers have directed their attention elsewhere, perhaps to other countries' data and other countries' problems. As a result, not a great deal is known about the effectiveness of our 'make work pay' policy. (Gregory 2013, p. 3)

The OECD has similarly drawn attention to a failure to provide data or conduct external evaluations of Job Services Australia (formerly the Job Network), casting Australia 'as secretive, relative to other countries' (OECD 2012, p. 225).

Sometimes government departments draw on administrative data but keep the evaluations in-house (McLachlan et al., 2013). Sometimes they will use outside researchers. The Department of Education, Employment and Workplace Relations has made unit record data available to the Melbourne Institute of Applied Economic and Social Research under a research agreement. This enabled analysis of the behavioural responses of income support recipients to a tightening in eligibility requirements in 2007 (Fok and McVicar 2011) and of their participation in training and education (Cai, Kuehnle and Tseng 2010). Arrangements such as this, while positive, are not broad enough and tend to be driven by the needs of government agencies, rather than releasing data per se for wider evaluation and analysis.

And we could do much more

At the state level, Western Australia (WA) has been an early adopter of making its state-based administrative data available. WA now has significant capability with data linkage and periodically has been able to access and link to Commonwealth data — typically for medical research — on a one-off basis after a protracted process. The statistical power of data linkage exercises and the consequent information made available for policy purposes are substantial (box 1.1).

Box 1.1 **The power of data linkage**

Medicines and birth defects

WA researchers linked PBS data with population-based data for over 100 000 pregnant women in WA from 2002 to 2005. Records of births to women who were dispensed medicines were linked to the Birth Defects Registry of WA. There were 47 medicines dispensed at least once during pregnancy with 23 associated with a registered birth defect to a woman dispensed the medicine. The study concluded that linked administrative data could be an important means of pharmacovigilance in pregnancy in Australia (Colvin et al., 2010).

Cancer risk from exposure to computed tomography (CT) scans

This study, funded by the Australian Government via the National Health and Medical Research Council, sought to assess the cancer risk in children and adolescents after exposure to CT scans. It covered 10.9 million people from Medicare records, aged 0-19 years in January 1985 and all Medicare-funded CT scans during 1985-2005 were identified. Diagnosed cancers were obtained from national cancer records. 60 674 cancers were recorded, including 3150 in 680 211 people exposed to a CT scan at least one year before any cancer diagnosis. Overall cancer incidence was 24 per cent greater than for unexposed people. The study concluded that future CT scans should be limited to situations where there is a definite clinical indication, with scans optimised to provide an image at the lowest possible radiation dose. (Mathews et al., 2013)

High care costs for mature aged Australians

A study undertaken by the University of Technology in Sydney examined health care costs for mature aged Australians by isolating expenditures due to health 'shocks' from those that are intrinsic to individuals. 267 000 survey responses obtained from the '45 and Up' study by the Sax Institute were linked to records from NSW Admitted Patient Data, NSW Emergency Department Data, the MBS and PBS. The NSW data linking was performed by the Centre for Health Record Linkage (CHeReL). The study found:

- high health expenditures that are intrinsic to individuals (or high fixed effects) tend to be associated with people who are old, sick and engage in unhealthy lifestyles.
- little evidence of high fixed effects being related to a relationship driven by a general practitioner nor by fee setting behaviour (Ellis et al., 2012).

Characteristics of children and families with child maltreatment

WA researchers investigated specific child and parental factors associated with increased vulnerability to substantiated child maltreatment. The study of all children born in WA during 1990–2005 used de-identified record linked data for child protection, disability services and health. The strongest factors found to increase the risk of child maltreatment included: children with an intellectual disability; parental socioeconomic status; parental age; and parental hospital admissions related to mental health, substance abuse and assault (O'Donnell et al., 2010).

Why isn't more happening?

Australia lacks a culture of information sharing and proactive data release. It appears that the main barriers to changing this culture are: protection of privacy; the resources needed to ensure that data are of sufficient quality for policy evaluation; and concerns by governments about unfavourable findings on policy effectiveness.

Privacy and confidentiality

Government agencies must ensure that personal information is not released publicly, is only available to authorised people on a need to know basis, cannot be derived from disseminated data, and is maintained securely. Linking administrative data or allowing access to third parties opens up further layers of risk, including attacks on data systems, either from within organisations, data laboratories, or through the internet (if accessible in this way).

Protocols for managing risks *ex ante* coupled with sanctions for researchers and data processors who breach privacy legislation are critical to assuage privacy concerns. Processes and systems can be implemented throughout data acquisition, storage and transformation to ensure data are secure, anonymous and accessed only by authorised individuals. Apart from standard de-identification protocols — regularly used by the ABS for example — more stringent safeguards can be implemented (box 1.2). Although some of these measures can reduce data quality somewhat, this is preferable to not releasing data at all.

De-identification of data, including setting up unique identifiers for matching, and storing these separately and securely, is feasible and commonplace. In relation to the WA data linkage system, Professor Stanley reported:

We've got registers of birth defects, of cancer ... of autism and mental health problems. We've got all the hospitalisations and all the deaths, and we collect these and link them together anonymously so that we actually only ever see the linked data. We're not interested in individual people; we're interested in large numbers ... (Stanley 2013)

In over 30 years of data linkage, the WA arrangements have not had one breach of any identifiable information (Stanley 2010, p. 75).

It is also notable that clients of services become frustrated when they have to submit the same information to different agencies because of privacy restrictions. Indeed, it appears that consumers in WA have lobbied for data linkage so as to improve services provision. A balance needs to be struck between information sharing and privacy by making clear that the purpose of using administrative data for research purposes is to benefit people, not to penalise them — fraud detection aside.

Box 1.2 Techniques to protect confidentiality

1. Suppression — not release parts of data that consist of too few observations.
2. Aggregation — make the data less precise by changing the level of detail.
3. Top/bottom-coding — limit the largest or smallest values possible of given variables.
4. Swapping — switch data values between records to make matching more difficult.
5. Random noise — add random amounts to numerical data, to mask the true amount.
6. Synthesising — replace data with values generated from probability distributions. Synthetic data can replace some variables or the whole data set (fully synthetic).

Sources: McCallister et al. (2010), Matthews and Harel (2011).

Resource implications and data quality

For administrative data to be useful for research it generally must first be manipulated (table 1.1). Data linking and matching can be complex especially where there are no unique identifiers. Automated matching and processing techniques can make linking data easier but these processes still require verification.

Researchers will want administrative data that is well specified, uses consistent definitions, and has ‘health warnings’ about pitfalls that might be known only to data owners. Even within series, discernible trends or deviations may simply reflect changes in definition. Databases need to be maintained and policy changes mapped. Clearly, there are non-trivial costs associated with maintaining, (dis)aggregating, linking, storing and supplying data. All of this requires specialist expertise, infrastructure and management time. Efficient user charges may be appropriate.

It would also be possible to reduce costs by anticipating data sharing. Greater prior consideration of the potential usefulness of data for research and evaluation could encourage more focused data collection, improving the quality of information for governments and reducing the reporting burden on providers. In its review into the *Contribution of the Not-for-Profit Sector*, the Commission found that agencies collected huge amounts of data from service providers, much of which was not used (PC 2010c). More useful data for providers would help them assess their own programs’ effectiveness, including through benchmarking against other providers. As observed by the Director of the Australian Institute of Health and Welfare (AIHW) in relation to ensuring the value of data sourced administratively:

One approach is to deliver some benefits to the provider of the information, so they not only incur the cost and inconvenience of the data supply, but also get some meaningful information back that helps them or their organisation to better carry out their required activities. (Kalisch, 2011, p. 7)

Greater use of data matching should encourage agencies to collect information in standard formats (e.g. the ranges used to collect income) which would increase the value of all existing data sets. Data matching could also reduce respondent burden by avoiding the need for repeated provision of the same information. The national information agreements signed up to by all governments and certain data providers, including the AIHW and ABS, should assist to improve the quality of administrative data for health, community services and housing. While these principle-based agreements are not binding, they can encourage better practice.

Political resolve

There is genuine appreciation by some data custodian officials of the power of administrative data. However, experience to date suggests that this appreciation has not been matched by improved access to that data for independent analysis. It appears that the blockages occur within policy departments, reflecting sensitivities that providing data for independent research could yield unfavourable public findings about policy effectiveness. Related to this is trepidation about releasing unrefined data and the misinterpretation or misuse of these data that could arise.

However, this short-term wariness comes at the cost of long-term gains for the Australian community. As noted, some Ministers have been more willing to allow researchers access to data, including the former Federal Minister for Human Services, who ‘swiftly approved data requests from RMIT University, the Australian National University and the University of Queensland’ (Martin 2012).

Other countries have shown resolve

Australia can look overseas to judge the feasibility and value of granting access to administrative data. In Denmark, Sweden, Finland and the Netherlands, linked administrative data are accessible for research purposes (Administrative Data Taskforce 2012). Statistics Finland considers that statistics should be compiled from administrative records whenever possible — around 96 per cent of its data come from these sources (Statistics Finland 2004).³ This openness promotes research — ‘microsimulation specialists pour into Nordic countries because of their liberal approach towards sharing statistics’ (Gruen and Goldbloom 2008).

In New Zealand education, migration, participation, social benefits and longitudinal business databases have been linked enabling research into areas such as: immigrant

³ Records include population, tax, trade, employment, labour market training, income support, conscription, student enrolments and business registrations.

outcomes; employment assistance effectiveness; effects of wage subsidies on individuals and firms; and intellectual property and productivity (Statistics New Zealand 2012, 2013). The New Zealand Government recently launched a system to give approved researchers remote access to de-identified microdata about people, households and businesses from their own desktops. The Minister for Statistics stated that the initiative was part of a ‘Government objective to have all public sector agencies releasing high value public data for re-use’ (Williamson 2013).

In Canada, administrative data on hospital discharges, prescription drug usage and ambulatory care is linked to population health survey data, birth and death databases and cancer registries (Statistics Canada 2010).

Australia — limited progress from sporadic starts

Western Australia’s Data Linkage System is seen by international peers as a leader in the field. Over 700 studies have drawn on the linked data in areas including health and aged care (formerly with the Commonwealth Department of Health and Ageing), development pathways for children, family connections, Indigenous identification, and road safety (DLWA 2013).

Progress in other Australian jurisdictions has been patchy. The Centre for Health Record Linkage, established in 2006, enables access to health data in New South Wales and the ACT (see box 1.1). It is one of the largest linked, health-related databases in Australia (CHeReL 2013). Queensland has recently made some databases available online and some other jurisdictions are making progress. The Queensland Premier stated that:

As a government, we collect, generate and use a lot of data. This data can deliver real benefits to the Queensland community and economy—if it is used in clever ways ... we will be releasing as much of it as possible ... (Queensland Government 2013)

Nationally, in 2008, Australian governments (through CoAG) agreed to make more administrative data available for performance reporting on health and education systems; disability, community and housing services; and the ‘Closing the Gap’ targets for Aboriginal and Torres Strait Islander Australians.

The Australian Government is in the early stages of developing a big data strategy to ‘enhance cross-agency data analytic capability for improved policy and service delivery’ (Commonwealth of Australia 2013, p. 4). Its issues paper highlighted the opportunities and challenges (e.g. privacy, data management and skills).

Drawing on the data linkage experience of WA, the Population Health Research Network (PHRN) is an Australian Government initiative to build a nationwide data

linkage infrastructure and enhance the way health and health related data are made available to approved researchers. It is a collaboration between the WA Centre for Data Linkage, Telethon Institute for Child Health Research WA, AIHW, the Sax Institute and state and territory data collation units. The PHRN *Proof of Concept Collaboration #1* project aims to link hospital admission data with hospital-related deaths across different states. The project will test data transfer and linkage processes. While most states have made progress developing linkage capabilities, lengthy delays occur with access to data owing to protracted approvals processes.

A Statistical Data Integration Involving Commonwealth Data (SDIICD) initiative was established in 2009 to ‘create an Australian Government approach to facilitate linkage of social, economic and environmental data for statistical and research purposes’ (CPSIC 2010, p. 2). A cross portfolio board oversees the data integration environment. All data integration projects under the SDIICD require an ‘Integrating Authority’ to be accountable for the project and projects considered high risk must use an ‘Accredited Integrating Authority’. The ABS and the AIHW are currently the only two accredited authorities (NSS 2011).⁴

While these institutional arrangements now in place could facilitate data linkage and access for research, it is important that they do not become too onerous and ‘chill’, rather than encourage, collaboration. For example, through its National Performance Reporting role, the Commission has found the SDIICD initiative requirements — such as the need to use a registered integrating authority rather than allowing work to be done in-house — to be unduly burdensome. In addition, while Ministers agree to the contents of National Minimum Data Set collections, which are managed by the AIHW, they insist on signing off any release of that data. The Commission has also asked the ABS to release non-contentious data under embargo for National Performance Reporting — as other data providers do routinely — but no action has occurred to date.

A sustained and concerted effort is needed

Policy-making based on good evidence is central to improving community living standards. Tackling community concerns about policy problems with expenditure announcements is not, of itself, sufficient. For expenditures to be effective and efficient they need to be based on analysis using the best information available. A rich vein of evidence resides within administrative databases. A failure to exploit

⁴ There are four projects on the Public Register of Data Integration Projects: ABS Census Data Enhancement Indigenous Mortality Project; ABS Migrant Personal Income Tax Data Integration Project – Feasibility phase; ABS Migrants Census Data Enhancement Project; Low dose radiation – effects of CT scans in childhood (AIHW).

this evidence would be a missed opportunity given Australia’s demographic and structural budget challenges.

The Australian Government has made statements recognising the benefits from better use of administrative data and introduced strategies and integration initiatives with new administrative architecture. All of this seems positive, but it has not yet been matched by open access to data for independent policy research. The frustrations here are eerily similar to those in the United Kingdom.

... there are examples in the UK of administrative data being linked between government departments and used for research purposes. However, the number of examples is too few, the time taken to get agreement to use such data is too long, inconsistent decisions are being taken within government departments concerning rules of access and, most frustratingly, the legislative framework provided to allow for linkages to be made across departments is cumbersome and inefficient. (Boyle 2012, p. ii)

There appears to be a similar lack of durable commitment by the Australian Government and most state and territory governments to make better use of data. On occasion, ‘reform champions’ within government have sought to release data in order to improve outcomes for the community, but sustaining momentum with changing personnel and shifting priorities is challenging.

Other nations and Western Australia — especially where it has been able to link to Commonwealth health data — have shown that harnessing administrative data can deliver substantial benefits with low risks, manageable costs and in ways that protect people’s privacy. Given the magnitude of current (and projected) expenditures in social programs, the relatively small costs of establishing systems for greater access to public data would be worthwhile.

Australia has an opportunity to support more open government, improve policy evaluation and strengthen public research. Realising these goals requires political will, articulated at the highest levels, to persevere with a concerted strategy with clear timeframes based on the principle that open access to de-identified information should be a default position. Realistically, it could take 5-10 years to rollout and embed systems before the ‘holy grail’ of relatively unimpeded remote access to high quality, de-identified and linked administrative data is achievable.

While there have been announcements and initiatives in the past and more recently, the lack of sustained tangible progress means that it is important that the 5-10 year timeframe does not become a motivation for more ‘false starts’, deferrals or eventual reprioritisation and non-delivery. International practices and over thirty years of experience in Western Australia suggest that the capabilities necessary to achieve a more open data culture could be developed by all Australian governments.