Chapter 5

Issues and barriers for researchers

Introduction

5.1   During the course of this inquiry, the committee was alerted to the many imperfections in Australia's data linkage system. This chapter examines a number of the most commonly raised issues and makes a number of recommendations aimed at progressing public health research in Australia.

5.2   This chapter examines issues that researchers and others may encounter in the process of attempting to obtain data held by government agencies. In particular this chapter will consider custodianship issues, the approvals process, linkage and the desultory departmental approach.

5.3   Australia's health data holdings are currently dispersed. The dispersed nature of the data means that the issues that follow are not a problem in all parts of Australia or are necessarily problems for all agencies. They are however, the issues that have been continuously raised with the committee.

5.4   Dr Merran Smith, Chief Executive of the PHRN conveyed the sentiment of many of the researchers, musing:

> …some of the problems with the Commonwealth that we are facing at this point in time is that, in my experience over many years, the Commonwealth government agencies often do not look outside. They are busy doing what they need to do in Canberra. Perhaps they do not appreciate, or are not in a position to leverage, the value and benefit that can be obtained from working with other groups within the nation.[1]

5.5   Certainly there is a view in the research community that 'departments tend to want to own that information and control it and use it for their purposes as opposed to a public purpose'.[2]

5.6   Ms Helen Owens, an Assistant Secretary at the Department of the Prime Minister and Cabinet, told the committee that the Prime Minister moved data policy into his department to try and improve the coordination of government data policy.[3]

5.7   As part of that effort, the Department of the Prime Minister and Cabinet have produced a *Public Sector Data Management report* and a *Public Data Policy*

---

1   Dr Merran Smith, Chief Executive, PHRN, *Committee Hansard*, 11 December 2015, p. 26.

2   Prof Brendon Kearney, Chair, PHRN, *Committee Hansard*, 11 December 2015, p. 28.

3   Ms Helen Owens, Assistant Secretary, Department of the Prime Minister and Cabinet, *Committee Hansard*, 3 February 2016, p. 13.

Statement. However, as Professor Kearney, Chair of the PHRN identified 'it is a long and slow process and a lot more needs to be done'.[4]

Custodianship issues

Accessing data

5.8   The first step in any data linkage project is to determine whether the required data is available and what permissions are required to access it. That process can be more difficult than one might anticipate.

5.9   Dr Heather Gidding, a Senior Lecturer in the School of Public Health and Community Medicine at the University of New South Wales observed that in undertaking her data linkage project:

> It took us a long time to find the data custodian for that ACIR [Australian Childhood Immunisation Register] data. When we did find someone who did understand the dataset, they are very hard to access. They are overworked and hard to get a hold of. It certainly does make the research quite difficult.[5]

5.10   The frustration and uncertainty about which department holds particular data was explained by Professor John Glover, Director of the Public Health Information Development Unit (PHIDU):

> One of the difficulties with the datasets we cannot get is…who owns which data, who the data custodians are. We go to the Department of Human Services and put a case to them for health data. They have a committee. Health [the Department of Health] sits on that committee and then basically decides whether or not you can have the data. The DHS people do not really know anything about the data and the policy issues; that comes from Health, and that decision is then made and passed back by email. You then complain and you make some comment, 'What about so and so, can we do this or that?, and then it goes back again. The same with PBS.

> It is very hard to engage totally with whoever the custodian really is—in this case it is Health, even though you are dealing with the Department of Human Services, who would run the tables.[6]

5.11   If researchers can work out which department holds the relevant data, Dr Smith explained that there is no clear process to access data:

> Part of the challenge with accessing the Commonwealth data is that there is not an explicit process. When you are accessing state data, most states have online an application form and an explanation for how you need to go about it. But even now when you go online and try to find how to access data,

4   Prof Brendon Kearney, Chair, PHRN, *Committee Hansard*, 11 December 2015, p. 28.

5   Dr Heather Gidding, Senior Lecturer and NHMRC Early Career Researcher, School of Public Health and Community Medicine, University of New South Wales, *Committee Hansard*, 11 December 2015, p. 36.

6   Professor John Glover, Director, PHIDU, *Committee Hansard*, 3 February 2016, p. 11.

there is a little bit on the AIHW site, hardly anything useful on the Department of Health's site or on the Department of Human Services' site. To acknowledge the work of my colleagues in the Commonwealth, they certainly did as a result of this process go through some conversations and streamline processes, but it is still not really explicit in the way that a researcher who says, 'Look, I think I want to use this Commonwealth data set' needs. It is not really clear how they go about the process to get the approval.[7]

5.12   Part of the problem is that Australian Government departments are often unaware of what data they hold and no list is publicly available.

5.13   This problem was highlighted when Ms Michelle Wilson, a General Manager at the Department of Human Services was unable to tell the committee what data the Department of Human Services held.[8] However, the Department of Human Services were not alone. The committee placed questions on notice to obtain a list of the health related data sets held by several health-focussed Australian Government departments and agencies.[9] Some departments required more than two weeks to compile their lists. The AIHW provides a list in its annual report.[10] The lists provided by the departments appear in Appendix 4.

5.14   Adding to the complication facing researchers of discovering what data sets are held by each department, some departments hold data that are owned by other departments. Ms Wilson explained:

> The Department of Social Services are the data custodians of [Centrelink] data, so while the Department of Human Services holds [the Centrelink] data, we are not the custodians of it. It is the same as with the Medicare MBS and PBS data that we collect in the course of our business: we are not the custodians of that data; we are the holders of it.[11]

5.15   Insofar as the Department of Human Services holds data on behalf of other departments, it considers itself to be a 'steward' of those data. Ms Wilson explained that a data 'steward', is 'somebody who sees the data through and ensures that it is well looked after and that the guidelines are adhered to'.[12]

---

7   Dr Merran Smith, Chief Executive, PHRN, *Committee Hansard*, 11 December 2015, p. 30.

8   Ms Michelle Wilson, General Manager, Department of Human Services, *Committee Hansard*, 11 December 2015, p. 69.

9   Department of Health, *Answer to Question on Notice 59*, pp 2–3; Department of Social Services, *Answer to Question on Notice 55*, p. 17; Australian Bureau of Statistics, *Answer to Question on Notice 54*, p. 1.

10   AIHW, *Annual Report 2014-15*, pp 201–204.

11   Ms Michelle Wilson, General Manager, Department of Human Services, *Committee Hansard*, 11 December 2015, p. 69.

12   Ms Michelle Wilson, General Manager, Department of Human Services, *Committee Hansard*, 11 December 2015, p. 69.

5.16   If researchers manage to navigate the challenges in finding the dataset and the custodian and obtain the relevant permissions, it is possible that the department may not provide the data in the level of detail required for the project. In this regard, Professor Glover explained to the committee:

> …a major gap in the data we have relates to the Medicare Benefits Schedule data and the Pharmaceutical Benefits Scheme data. We cannot get it at what I believe is an appropriate geographic level—a level that will show the major variations that occur within our cities and within our regional and rural areas, particularly in variations in health outcomes, health service youth and social determinants of health.[13]

Under-resourced

5.17   The committee heard from multiple witnesses that the data custodians they dealt with were not sufficiently resourced to ensure that the data asset could be maximised. For example Professor Sallie-Anne Pearson, Head of the Medicines Policy Research Unit, Centre for Big Data Research in Health acknowledged that:

> One of the greatest challenges [custodians] face is that this is incredibly under-resourced, and designated roles and having people responsible to do this work would be a very pragmatic and significant advance in getting some of those current logjams freed.[14]

5.18   Professor Louisa Jorm, Director of the Centre for Big Data Research in Health echoed those comments:

> I would second that. In fact, several data custodians have said to me over the years: 'Providing data for research is not core business for our agency.' It is viewed as an add-on… If the leaders of their agencies are saying, 'This is your core business and we're going to resource it,' then things will start to move.[15]

5.19   The committee attempted to gauge whether recent shifts in government priorities may mean that appropriate resources will now be allocated to data custodians.

5.20   Ms Foster, First Assistant Secretary with the Department of Health explained to the committee some of the evolving processes within the Department of Health:

> In terms of access to data and better use of data, the Department of Health has been through a number of different reviews that talked about better use of data, better valuation, better research to inform policy development towards better health outcomes. In fact, that was the reason the division I head up was established. It was to try and make better use of data and

---

13   Prof John Glover, Director, PHIDU, *Committee Hansard*, 3 February 2016, p. 2.

14   Professor Sallie-Anne Pearson, Head, Medicines Policy Research Unit, Centre for Big Data Research in Health, *Committee Hansard*, 11 December 2015, p. 12.

15   Professor Louisa Jorm, Director, Centre for Big Data Research in Health, *Committee Hansard*, 11 December 2015, p. 12.

ensure better use of evaluation. Certainly, in my KPIs it would be a fairly key feature that there be better use of data. In the department's vision…there is that reference to better health outcomes and one of the key underpinnings of that is that better use of research and data. So it really is quite widespread throughout the department. There is a major cultural change that is occurring.[16]

5.21   However that statement was not reflected in the Department of Health's Answer to a Question on Notice. When asked about the KPIs the Department of Health uses to evaluate data usage the department provided the following cryptic response:

The Department facilitates a Data Governance Council that includes representation from the Department, the AIHW, the Australian Bureau of Statistics, Department of Human Services and other Health portfolio agencies. The Council is responsible for ensuring effective policies and governance for the Department's approaches to data collection, management, interrogation, sharing, access and release.[17]

5.22   The Australian Bureau of Statistics identified that it has a KPI entitled 'Use of Microdata Increases'.[18] The AIHW has measures relating to data tied into its Portfolio Budget Statements.[19] However, it appears that neither the Department of Health, the Department of Human Services or the Department of Social Services have any KPIs relating to data.[20]

The changing data culture

5.23   Professor Pearson pointed out that data custodians play an important role in releasing health-related data:

The custodianship of the data is a very important position to be placed in. I think traditionally custodians have been very risk averse. There are different interpretations of legislation guidelines, et cetera, about the nature and the type of data that can be released. I have actually seen a massive change over the last few years where custodians are now thinking about arguments as to why data should not be released…rather than thinking, 'We have to stop this.'[21]

---

16   Ms Alanna Foster, First Assistant Secretary, Department of Health, *Committee Hansard*, 11 December 2015, p. 73.

17   Department of Health, *Answer to Question on Notice 61*, p. 7.

18   Australian Bureau of Statistics, *Answer to Question on Notice 53*, p. 6.

19   AIHW, *Answer to Question on Notice 56*, p. 9; AIHW, *Annual Report 2014-15*, www.aihw.gov.au/aihw-annual-report-2014-15/ch1/ (accessed 9 February 2016).

20   Department of Social Services, *Answer to Question on Notice 55*, p. 15; Department of Human Services, *Answer to Question on Notice 63*, p. 14.

21   Professor Sallie-Anne Pearson, Head, Medicines Policy Research Unit, Centre for Big Data Research in Health, *Committee Hansard*, 11 December 2015, pp 9–10.

5.24   Professor Pearson also noted:

> There clearly is an enormous responsibility on the part of custodians to protect the identity of individuals, but the challenges that they face relate to the interpretation of legislation, which can be different with different individuals. The other great challenge that they face is that oftentimes this role is done on top of a highly powered, high-level and highly committed role. It is really about goodwill and trying to serve the research community.[22]

5.25   Apparently a new culture, with regard to the use and release of administrative data held by Australian Government departments, is in its infancy. While change is occurring, the speed at which parts of the public service are adapting varies. An example of the more cautious approach was exemplified by Ms Foster of the Department of Health:

> …the whole senior executive [are] helping to drive that cultural change to make data available.

> But having said that, while we are moving from that attempt to make data as available as possible to researchers and for policy development purposes there is that very key issue of privacy that we are concerned about. I think that for many years there was that culture, 'We must absolutely protect this data at all costs.' But, of course, as techniques—computing and statistical techniques and all sorts of things—get more sophisticated there are more ways to 'perturb' the data, as I think you referred to it, or to confidentialise the data so we can actually protect people's privacy and still be able to make information available for use by researchers.[23]

5.26   Ms Foster continued:

> There are very strict guidelines under the National Health Act, the Health Insurance Act, the privacy guidelines and the Privacy Act. We also observe those provisions very strictly. Indeed, sometimes those rules can limit our own potential to use data internally. But we do have to be very conscious of those privacy rules to ensure that people's information cannot be identified.[24]

5.27   Ms Foster's answer can be starkly contrasted with the dynamic response of Mr David Dennis, a Branch Manager with the Department of Social Services who submitted:

> We have endeavoured, especially in recent times, to be at the vanguard of testing technologies and new techniques and developing new techniques to

---

22   Professor Sallie-Anne Pearson, Head, Medicines Policy Research Unit, Centre for Big Data Research in Health, *Committee Hansard*, 11 December 2015, p. 12.

23   Ms Alanna Foster, First Assistant Secretary, Department of Health, *Committee Hansard*, 3 February 2016, p. 23.

24   Ms Alanna Foster, First Assistant Secretary, Department of Health, *Committee Hansard*, 3 February 2016, p. 23.

allow more data to be released to intelligent sources... Internally, we have worked with DHS to streamline our approval processes and, in so doing, have devolved a lot of routine approvals to DHS directly so that they no longer need to seek recourse to DSS where routine approvals are involved, so it moves a lot more quickly.

Where we have been successful in this regard is by not treating every data application equally. In so doing, we have published to data.gov.au large amounts of information that is now open, albeit in a perturbed or confidentialised manner. That has allowed us to turn our minds to what we see as the current challenge…of arranging for researchers to have access to individual unit record data in a safe and appropriate environment, because this is what they need to perform the statistical analyses that they want to perform. We have trialled a number of models to facilitate that…using innovative products…such as the Secure Unified Research Environment… so that finally we can put an enduring dataset of DSS data in a safe place that credentialled individuals can access directly at the unit record level and where they can withdraw the products of those analyses, and we are assured that the data is safe and the researchers get the access that they need...[25]

5.28   The Australian Bureau of Statistics also appears to be embracing the new data culture. The Australian Bureau of Statistics now operates a number of linked data projects including the Mental Health Services-Census Integrated Dataset.

5.29   The Mental Health Services-Census Integrated Dataset is a good example of linking diverse data sets to provide a solid evidence base. It combines 2011 Census data with mental health related items from the MBS and PBS along with demographic data. The Australian Bureau of Statistics is currently investigating adding state and territory hospital data to this dataset to enable 'a more complete picture of the patterns of service usage'.[26]

Approvals processes

5.30   Various approvals to access data and conduct research are necessary to ensure that the data being released are used for the public benefit and that appropriate measures are being taken to ensure the privacy and security of personal information.

Ethics approvals

5.31   To ensure that health information is used appropriately, delegated legislation provides that researchers must undertake ethics approval if they wish to use health data.[27]

---

25   Mr David Dennis, Branch Manager, Department of Social Services, *Committee Hansard*, 3 February 2016, p. 25.

26   Australian Bureau of Statistics, *Submission 192*, p. 4.

27   See *Privacy Act 1988*, ss 16B, 95, 95A and *Guidelines under Section 95 of the Privacy Act 1988*.

5.32   Ethics approvals play an important part in upholding high ethical standards in research. However, as Dr Gidding told the committee, having to satisfy multiple approvals can add substantial time to a research project:

> I am leading a team of researchers in New South Wales and Western Australia that predominantly have linked up the Australian Childhood Immunisation Register, the vaccination records of Australian children to records for children in both WA and New South Wales—1.8 million children… We had to get nine approvals for our project to go ahead, which also included six ethics applications that took two years.[28]

5.33   The Australian Commission on Safety and Quality in Health Care (ACSQHC) endorsed the idea that multiple ethics approvals hindered data linkage:

> One of the main issues is…[the] complicated ethics approvals systems in the various jurisdictions. In Australia, projects using linked data must be approved by the data linkage unit; the data custodian responsible for each data set; and one or more Human Research Ethics Committee(s)… In order for data on safety and quality to be most effective in driving improvement, regular and timely review is required.[29]

5.34   Dr Nicky Antonius, Acting Assistant Secretary of the Department of Health and Mr Warren Richter, Head, Chief Information Officer Group, AIHW both told the committee they would not request an additional ethics approval where the researcher had already obtained an approval that met the research ethics standards of the National Health and Medical Research Council (NHMRC). However, ethics approvals from other jurisdictions were more problematic. This continues to be a problem even within Australia where Commonwealth departments will not accept the ethics approval of a State, such as Victoria.[30]

5.35   In his evidence to the committee, the Acting Australian Information Commissioner Mr Timothy Pilgrim questioned whether ethics approvals were necessarily the best mode of protecting privacy:

> I have heard researchers remark regularly about some of the challenges around ethics committees, and there may be a good argument for looking at other mechanisms.[31]

---

28   Dr Heather Gidding, Senior Lecturer and Early Career Researcher, University of New South Wales, *Committee Hansard*, 11 December 2015, p. 36.

29   ACSQHC, *Submission 187*, pp 2–3.

30   Dr Nicky Antonius, Acting Assistant Secretary, Department of Health and Mr Warren Richter, Head, Chief Information Officer Group, AIHW, *Committee Hansard*, 11 December 2015, p. 72.

31   Mr Timothy Pilgrim, A/g Australian Information Commissioner, *Committee Hansard*, 2 February 2016, p. 11.

Delays

5.36   Once a researcher has obtained ethics approval, the committee received evidence that there have often been significant delays in obtaining data from departments.

5.37   In 2009 Professor Fiona Stanley and Emeritus Professor D'Arcy Holman wrote a letter to the then Secretary of the Department of Health and Ageing, Ms Jane Halton. The letter notified the Secretary:

> Of the 23 research projects [that had applied for Commonwealth held health data], only nine have received useable Commonwealth data. All nine have experienced lengthy delays in obtaining data, some as much as two years from the time of application. Those still to receive data have been waiting for as long as three years. It would be fair to say that the affected researchers have found it extremely difficult to engage with the Australian Department of Health and Ageing in progressing a solution to these delays and solving attendant issues such as data queries and explanations.[32]

5.38   Similarly when Professor David Preen, Deputy Head of the School of Population Health at the University of Western Australia, was asked about the source of delays in cross-jurisdictional linkage, he responded:

> The approval processes through the Commonwealth were slow... [In] a number of cases we had queries or had made submissions but did not hear back for a period of time, and we would follow up. But it was not clear to whom we should be following up with... But, there being delays, what we could do to facilitate that and what else may needed to provide in terms of information, et cetera, was very unclear. That, in itself, caused further delays. Whether that was due to inaction or a lack of transparency with the existing processes, I guess, is debatable. But the end result is a huge delay in provision of these data.[33]

5.39   It is important to note that there is a large financial cost to researchers from these delays. In particular, the lengthy delays have led to a significant wastage of public research funds.[34]

5.40   The degree of wastage was articulated by Emeritus Professor D'Arcy Holman who estimated that in 2008:

> …over 100 medical researchers waiting for [Commonwealth] data, had $11 million in precious public research funds, mostly from NHMRC, simply going to waste.[35]

---

32   Prof Fiona Stanley and E/Prof D'Arcy Holman, *Tabled Document 98*, p. 2.

33   Prof David Preen, Deputy Head, School of Population Health, University of Western Australia, *Committee Hansard*, 11 December 2015, p. 41.

34   Telethon Kids Institute, *Data Linkage Roundtable: Summary and actions*, p. 1, http://telethonkids.org.au/media/1033902/Data-Linkage-Roundtable-Report.pdf (accessed 12 February 2016).

5.41   The problem of significant delays in obtaining data is not limited to researchers. It also affects government agencies. In its most recent annual report the AIHW explained the impact of the delays in producing its reports:

> This was because we did not gain access to records from the Medical Benefit Schedule and the Pharmaceutical Benefits Schedule as expected, which meant that the committee [the AIHW Ethics Committee] did not receive applications from external researchers for ethical clearance of projects involving use of these particular data sets.[36]

Linkage

Linkage logjam

5.42   The committee heard that one of the greatest impediments to the widespread use of linked data is the lack of facilities accredited to link Commonwealth data.

5.43   Professor Louisa Jorm described the current situation:

> …the bottleneck is our limited capacity for data linkage. Currently only the Australian Bureau of Statistics and the Australian Institute of Health and Welfare are accredited as integrating authorities for Commonwealth health data, and yet we have state-of-the-art health data linkage facilities operating in states and territories, including New South Wales, South Australia and the Northern Territory, and Western Australia. National capacity for data linkage would be dramatically boosted if these state and territory based units were accredited to link Commonwealth as well as jurisdictional data. In particular, it makes a lot of sense for these units to service projects that involve linkage of data from a single jurisdiction with Commonwealth data, rather than these projects adding to the linkage logjam and linkage queue that currently exists at the Australian Institute of Health and Welfare.[37]

5.44   Similarly, the Centre for Big Data Research in Health strongly advocated for State and Territory data linkage units to be accredited by the Commonwealth:

> Australia has world-renowned health data linkage facilities yet relatively few of these are accredited to link and integrate Commonwealth data collections. Health data linkage activities have increased substantially over the last decade and this activity will continue to grow in the coming years. The Commonwealth data linkage facilities based at the ABS and AIHW are not necessarily equipped to manage this growth and are not always an efficient option for projects involving linkage of Commonwealth data with

35   Emeritus Professor D'Arcy Holman, *Additional Information 13*, p. 14.

36   AIHW, *Annual Report 2014-15*, p. 6.

37   Professor Louisa Jorm, Director, Centre for Big Data Research in Health, *Committee Hansard*, 11 December 2015, p. 7. The Australian Institute of Family Studies is also an integrated authority but is yet to complete a data linkage project.

collections from one or more States. One solution is to let others with proven track records do this too.[38]

5.45   According to the National Statistics Service, the only Integrating Authorities accredited to link Commonwealth health data are the ABS, the AIHW and the Australian Institute of Family Studies.[39]

5.46   Dr Smith advised the committee that attempts to have State data linkage units accredited have been unsuccessful:

> One of the state linkagers has sought accreditation to be able to receive Commonwealth data and the view from the Commonwealth was that they could not accredit it because it was a state agency, so they did not have jurisdiction.[40]

Cost of linkage for researchers

5.47   The committee heard evidence that data linkage can be a costly barrier to the development of new knowledge and innovation.

5.48   Some Australian Government departments charge researchers on a cost recovery basis to access the data for linkage projects.[41] In some cases, the amount charged is cost-prohibitive. SA-NT Datalink provided two examples:

> For one project with a cohort of about 10,000 individuals and linking 4 datasets, the SA-NT DataLink cost was estimated at $10,000. The researchers also wished to link to a Commonwealth dataset for which they quoted approximately $160,000.

> Because of the high Commonwealth costs, the researchers could not include this data.

> Another project with a cohort of about 240 individuals and linking 4 datasets, the SA-NT DataLink cost was estimated at $8,500. The researchers also wished to link to another Commonwealth dataset for which they quoted approximately $40,000.

> Again, because of the high Commonwealth data costs, the researchers could not include this data.[42]

---

38   Centre for Big Data Research in Health, *Submission* 172, p. 3.

39   National Statistics Service, *Accredited Integrating Authorities*, www.nss.gov.au/nss/home.nsf/pages/Data%20Integration:%20Accredited%20Integrating%20Authorities (accessed 14 January 2016). Integrating Authorities are currently accredited by the Cross Portfolio Data Integration Oversight Board.

40   Dr Merran Smith, Chief Executive, PHRN, *Committee Hansard*, 11 December 2015, p. 32.

41   See for example AIHW, *Answer to Question on Notice 58*, p. 7; Department of Human Services, *Answer to Question on Notice 63*, pp 11–12.

42   SA-NT DataLink, *Answer to Question on Notice 65*, 2 February 2016, p. 1.

5.49   Then linkage units can also charge researchers depending on the work performed. Officials at the Australian Institute of Health and Welfare told the committee that they operate on a cost-recovery basis:

> As you may be aware, the Australian Institute of Health and Welfare receives about 30 per cent of its funding from appropriations, so 70 per cent of our revenue comes from the provision of goods and services to others. We run our data-integrating authority and data-linkage services on a cost-recovery basis. That is, essentially, the cost of a salary plus the overheads associated with running buildings… There are other organisations involved in the chain. If you have a state-linkage organisation, a Centre for Health Record Linkage in New South Wales, they are also running on a cost-recovery basis, so the researchers have to pay the costs that are incurred to deliver the service they are asking for.[43]

5.50   When the committee asked NACCHO [the National Aboriginal Community Controlled Health Organisation] what it cost to link data through the AIHW Dr Robert Starling replied 'it is in the tens of thousands of dollars—it is not 50 bucks here or there.'[44]

5.51   In response to a question on notice, the AIHW advised that the average cost of a linkage project in 2015 was $4414 inclusive of GST.[45]

5.52   The ABS notified the committee that the average fee of obtaining Custom Data Information was $1970 but that average cost of using Information Consultancies carried out by the Health subject matter area was $4645 inclusive of GST.[46]

5.53   Researchers may also be charged to use laboratories such as the SURE research facility to analyse data. Mr Wells, the Deputy Chief Executive Officer of The Sax Institute informed the committee:

> On the researcher side, we do charge researchers. But that charge does not reflect the full cost, because we have a subsidy through NCRIS funding. So we charge them a charge which is not the full cost of providing it. But, certainly, it is a very costly facility to run. I am not saying that is a bad thing, but it is costly. We are constantly looking at enhancing its efficiency et cetera, but there is very much a cost to operating and providing this service.[47]

---

43   Mr Warren Richter, Chief Information Officer, AIHW, *Committee Hansard*, 11 December 2015, p. 70.

44   Dr Robert Starling, Chief Information Officer, NACCHO, *Committee Hansard*, 11 December 2015, p. 56.

45   AIHW, *Answer to Question on Notice 51*, p. 2.

46   ABS, *Answer to Question on Notice 53*, p. 5. ABS have an arrangement with Australian universities that all but the most complex work is provided to researchers free of charge. Complex work is still charged on a cost recovery basis.

47   Mr Robert Wells, Deputy Chief Executive, The Sax Institute, *Committee Hansard*, 11 December 2015, pp 26–27.

5.54   One strategy that has the potential to bring down the cost of linkage is the creation of enduring linked datasets.

Enduring data linkage

5.55   At the end of a linkage project there is a legal requirement that researchers must destroy the links between the datasets as Mr Crettenden, Assistant Secretary, Department of Health explained:

> The Privacy Commissioner's guidelines…say that Medicare…and the [Department of Health] are only able to link the datasets under very specific circumstances and that the dataset that is used to create the linkage must be destroyed within one month of it being created.[48]

5.56   However that creates problems for ongoing research according to Dr Smith:

> If you go to a lot of effort to create a high-quality linkage and then, after the project is finished, you destroy the links and start doing the same thing again for the next project, first of all, it is highly inefficient; it is both expensive and time consuming. Second, it does not lead to improvements in linkage quality. When we talk about the importance of enduring linkage, that is why it is important.[49]

5.57   The benefits of being able to maintain a linked dataset is that the data can be accessed very quickly and it can become a permanent resource that researchers can use.[50]

5.58   Australia has created enduring datasets before as Professor Preen recalled:

> It was a formalised and accepted process for linkages de-identified unconsented or non-consented Medicare, PBS and state data. In WA a memorandum of understanding was developed and signed by the Commonwealth and by the state government to set up what was referred to as the cross-jurisdictional linkage facility to specifically link the projects after approval. These data would be linked in unconsented form but they would be provided in de-identified form for researchers. My understanding is that 25 projects went through, were approved and received data through that process. That system was functional, and the fact that the MOU was agreed to and signed means that it should be a possibility, obviously, within the existing legal framework. That has now been replaced because of some of the issues with the layers, with integrating authorities.[51]

---

48   Mr Ian Crettenden, Assistant Secretary, Department of Health, *Committee Hansard*, 11 December 2015, p. 67.

49   Dr Merran Smith, Chief Executive, PHRN, *Committee Hansard*, 11 December 2015, p. 23.

50   Mr Warren Richter, Group Head, Chief Information Officer Group, AIHW, *Committee Hansard*, 11 December 2015, p. 66.

51   Prof David Preen, Deputy Head of the School of Population Health, University of Western Australia, *Committee Hansard*, 11 December 2015, pp 42–43.

5.59   Departmental officials indicated that they were currently undertaking an ambitious Multi-agency Data Integration Project (MADIP). The MADIP aims to combine 'information from existing surveys, administrative collections and censuses' to develop 'a more complete picture of the circumstances of individuals, households and businesses'.[52] The MADIP 'brings together, for the first time, Census data with administrative data on health, income, and social security payments'.[53]

5.60   Unlike the 'link and destroy' model described above, MADIP is going to be an enduring linkage.[54]

5.61   Promisingly, there is a commitment amongst several key government agencies to move towards greater use of enduring data sets. Ms Gemma Van Halderen, the General Manager of the Strategic Partnerships and Projects Division at the Australian Bureau of Statistics told the committee:

> …the Commonwealth secretaries and some dep secretaries across the Australian Bureau of Statistics, Department of Human Services, Department of Social Services and the Department of Health have been working…since about 2009 on some Commonwealth arrangements for data sharing and data use. The secretary commissioned a review of those arrangements in 2014. That review was undertaken by a consultant, Ms Susan Linacre, who used to be a deputy at the Australian Bureau of Statistics. She recommended to the deputy secretaries and the secretaries that the Commonwealth should move away from a link-and-destroy model to a create-and-reuse-and-keep model, so we are currently moving towards that… There is a very strong appetite within the Commonwealth agencies and the members of that data integration initiative to put in place a research create-and-reuse-and-keep model, not a create-and-destroy model for data linking.[55]

The desultory departmental approach

5.62   As the above evidence reveals, there is no consistent policy between Australian Government departments in their approach to releasing data to researchers.

5.63   Part of the reason for the haphazard approach may be attributed to the lack of coordination of government policy in this area. Ms Owens, Assistant Secretary, Department of the Prime Minister and Cabinet informed the committee that her department is responsible for the coordination of this area of policy:

> Our role is really the coordination of the policy element. We would still rely upon our colleagues here at the table to do the actual coordination within

---

52   Australian Bureau of Statistics, *Submission 192*, p. 2.

53   Australian Bureau of Statistics, *Submission 192*, p. 2.

54   Dr Nicky Antonius, Acting Assistant Secretary, Department of Health, *Committee Hansard*, 3 February 2016, p. 17.

55   Ms Gemma Van Halderen, General Manager, Strategic Partnerships and Projects Division, Australian Bureau of Statistics, *Committee Hansard*, 3 February 2016, p. 22.

their own thematic areas. We are not actually going to do the doing in this space...[56]

5.64   The problems identified in this report highlight the failures of some departments to adapt to a new, more flexible approach to releasing de-identified data. Those failures must be addressed and rectified before Australia can make the most of its big data potential.

Release of information between government departments

5.65   The complexity of the Australian Government's data arrangements were reinforced when Senator Moore tried to ascertain what data was held by the Department of Human Services:

> **Senator MOORE**: I am interested to know whether the human services data includes only Medicare data or whether it has Centrelink and DVA [Department of Veterans' Affairs] data as well…
>
> **Ms Wilson**: …The Department of Social Services are the data custodians of that data, so while the Department of Human Services holds that data, we are not the custodians of it. It is the same as with the Medicare MBS and PBS data that we collect in the course of our business: we are not the custodians of that data; we are the holders of it.
>
> **Senator MOORE**: Are you custodians of anything?
>
> **Ms Wilson**: We are custodians of the data about how our customers behave in our service delivery systems.[57]

5.66   Dr Nicky Antonius, Acting Assistant Secretary at the Department of Health also tried to assist the committee explaining:

> If DHS [the Department of Human Services] was asked to divulge data to AIHW [Australian Institute of Health and Welfare] that would depend what dataset and who is the custodian of that dataset. For MBS, PBS and Medicare enrolments data it would be the Department of Health but we are the ones who will pursue the public interest certificates and then share that with the Department of Human Services to forward it to the requester. If the requester is asking for social services data then they would be directed to the Department of Social Services.[58]

5.67   Mr Warren Richter of the AIHW and Ms Michelle Wilson from the Department of Human Services told the committee that the process of releasing data between Australian Government department could be very involved:

---

56   Ms Helen Owens, Assistant Secretary, Department of the Prime Minister and Cabinet, *Committee Hansard*, 3 February 2016, p. 14.

57   Ms Michelle Wilson, General Manager, Department of Human Services, *Committee Hansard*, 11 December 2015, pp 68–69.

58   Dr Nicky Antonius, A/g Assistant Secretary, Department of Health, *Committee Hansard*, 11 December 2015, p. 69.

Mr Richter: We have a schedule agreement that was signed last week with the department [Department of Human Services]. We have a public interest certificate which is in the process of being signed within the department and we have arrangements in place with the Department of Human Services to receive the Medicare enrolments data, which we need to do the linkage—to identify the links. As soon as that public interest certificate has been signed—I do not know if you know, Michelle, but we have your people all on tap to hit the button—

Ms Wilson: We are looking at the piece of equipment that is holding the data to make sure it is ready to go. We will wait until the public interest certificate has been tied up with a bow and signed by everybody. It is going through the final processes, as is the multi-agency data integration project. The public interest certificate is also at that final stage of processing. Each of our departments has to go through its own legal processes.[59]

5.68   That somewhat confusing state of affairs was compounded by Ms Wilson who went on to argue:

We [the Department of Human Services] relied on a public interest certificate and a collaborative process to look at the privacy and secrecy issues with the Department of Health. Indeed, with the Department of Social Services data that relates to social security information we also have a collaborative process with the Department of Social Services. So it is generally not just up to the Department of Human Services about how the data is used. In fact we cannot really make decisions ourselves about how that data is used.[60]

5.69   The *Public Sector Data Management Report* provides additional examples of current inefficiencies:

In one case where an important dataset was being linked with the Census, the negotiations on the MOU [Memorandum of Understanding] took up to 18 months while the linking only took two weeks.[61]

5.70   And:

An agency reported having up to 11 MOUs with the same department to access data.[62]

5.71   Accordingly, the Public Sector Data Management Report concluded:

---

59   Mr Warren Richter, Chief Information Officer, AIHW and Ms Michelle Wilson, General Manager, Department of Human Services, *Committee Hansard*, 11 December 2015, p. 68.

60   Ms Michelle Wilson, General Manager, Department of Human Services, *Committee Hansard*, 3 February 2016, p. 24.

61   Department of the Prime Minister and Cabinet, *Public Sector Data Management Report*, December 2015, p. 18, www.dpmc.gov.au/pmc/publication/public-sector-data-management-report (accessed 5 January 2016).

62   Department of the Prime Minister and Cabinet, *Public Sector Data Management Report*, December 2015, p. 18.

Overall, the lack of data sharing prevents feedback on policy and hinders the potential of data to improve future service delivery.[63]

Committee view

5.72   The committee is deeply concerned by the lack of transparency regarding the data held by government departments. If Australia is to maximise the value of its data resources, researchers need to know what data sets are available and how to access them.

5.73   During the course of this inquiry it became obvious that some departments were uncertain about what datasets they held. In some cases, when asked to provide a list to the committee on notice some departments required two weeks to list their data holdings.

5.74   It ought to be clear to both researchers and departments what data each department collects and what area within the department is responsible for data custodianship. The *Public Data Policy Statement*, the *Public Sector Data Management Report* and a chorus of witnesses all agree that non-sensitive data should be publicly available as a right and that sensitive data should be able to be accessed subject to appropriate privacy and security constraints. For this purpose, a list of data sets held by each department ought to be publicly available on their websites and on data.gov.au. Departments also ought to publish a clear statement that identifies exactly how requests for data are made and how those requests will be dealt with.

Recommendation 6

5.75   The committee recommends that each Australian Government agency develop and maintain on its website a list of datasets held by the agency along with the contact details of the data custodian. This list should be updated at least twice annually.

Recommendation 7

5.76   The committee recommends that all datasets held by the Commonwealth be listed on www.data.gov.au, identifying which agency is the data custodian.

Recommendation 8

5.77   The committee recommends that each Australian Government agency that is a data custodian develop and publish on its website guidance for researchers detailing its process for data requests and approvals.

---

63   Department of the Prime Minister and Cabinet, *Public Sector Data Management Report*, December 2015, p. 19.

5.78   In certain departments the current system of approving access to data is at times chaotic. The evidence demonstrates that seeking multiple ethics approvals was time consuming without necessarily being of significant additional benefit.

5.79   The committee is of the view that there ought to be a consistent method of applying for approval to access sensitive data. The committee notes that ethics approvals of other states are currently not recognised by Australian Government departments. Such approvals ought to be nationally recognised to ensure consistent and systematic application of process. Therefore, the relevant departments need to work at both the Commonwealth level and with the States and Territories.

Recommendation 9

5.80   The committee recommends that the government take a whole-of-government approach to streamlining the ethics approval process and the authorising environment in consultation with the Privacy Commissioner, privacy advocates, the NHMRC, data custodians, academics, consumers and the States and Territories. The government should also work with the States and Territories to establish a national accreditation system so that ethics approvals from accredited jurisdictions are recognised by the Commonwealth.

5.81   The evidence received by the committee demonstrates that data custodians are currently not being adequately resourced. Furthermore, it was clear from the lack of KPIs in most government departments that the responsibilities of data custodianship is not a priority for departments despite the considerable benefits that would flow from a sharper focus.

5.82   The burdens placed on custodians have undoubtedly made them more hesitant to release data. The relevant departments need to empower the officials in these positions and give them the resources and confidence to be able to release datasets where researchers have obtained the approval that demonstrates their ability to maintain the security and privacy of the data.

5.83   The committee was shocked to discover that precious Commonwealth funded grants for important medical research projects were going to waste because Australian Government departments would not provide the data required. The committee believes that the $11 million figure cited by Emeritus Professor Holman is highly conservative.

5.84   The committee notes that the Productivity Commission and others have urged government to make de-identified datasets more readily available to researchers. Where data has been de-identified, open access ought to be the default position.

5.85   It is absolutely unacceptable that researchers should have to wait years to access data to facilitate important research projects in the public interest. This is a situation that requires immediate attention.

5.86   Therefore the committee makes the following recommendations aimed at prioritising data access and encouraging appropriate open access.

Recommendation 10

5.87   The committee recommends that relevant government agencies give greater priority to, and adequately resource, their data custodians.

Recommendation 11

5.88   The committee recommends that relevant government agencies provide guidance to data custodians to assist them in their decision-making, with a view to making more de-identified data available on an enduring basis.

Recommendation 12

5.89   The committee recommends that the government adopt the Productivity Commission's proposed principle that open access to de-identified datasets should be the default position.

Recommendation 13

5.90   The committee recommends that the government should direct relevant agencies to release de-identified datasets on an enduring basis as the default position.

Recommendation 14

5.91   The committee recommends that departments that have data custodianship responsibilities must establish and publish realistic Key Performance Indicators for the timely consideration and approval of datasets requests. These departments must publicly report on their KPIs in their annual reports.

If after 5 years departments continue to delay the release of datasets, then the committee recommends that the government establish binding timeframes for processing applications for data. Failure to comply with the timeframe should trigger appeal rights similar to those found in other information access regimes.

5.92   Currently, the release of various datasets between Australian Government departments and agencies at times appears shambolic. The evidence provided to the committee indicates that several departments, despite the imprimatur of the new data policy, were very protective of their data and demonstrated great reticence to release it to other government agencies. Whilst the committee understands and supports the need to protect privacy concerns and to act in accordance with legislation, the current practices are inefficient and stifle innovation.

Recommendation 15

5.93   The committee recommends that Government encourage collaboration on data linkage projects between government agencies, as well as academia and industry to provide for evidence-based policy development and facilitate research that is undertaken in the public interest.

5.94   The committee is confident that increasing the availability of linked data will make Australia a more attractive research destination that produces significant public health research and better evidence-based policy proposals.

5.95   To ensure that researchers can be supported into the future, consideration ought to be given to the linkage units. The committee heard that some units were almost already at capacity and that there was at the very least the potential for a 'linkage logjam'.

5.96   Australia has a number of excellent linkage units at the state level, some of whom would like to be accredited to link Australian Government data. The government should give the prospect serious consideration.

5.97   The committee also notes that depending on the nature of the project some researchers may be charged multiple times (potentially at the Commonwealth and state levels) for accessing various services in order to facilitate that research. To ensure that the medical research sector is as vibrant as it can be, the government ought to consider the cumulative costs incurred by researchers across the linkage project.

Recommendation 16

5.98   The committee recommends that government consider accrediting State data linkage units to link Commonwealth data with State data collections, subject to comprehensive privacy and security protocols.

Recommendation 17

5.99   The committee recommends that the Government review the cost of data access and linkage work undertaken by Commonwealth entities with a view to facilitating research and innovation in the national interest.

5.100   Australia has the capacity to be a world-leader in this field. For the benefit of the health of all Australians the committee strongly urges the government to implement all recommendations to ensure that Australia realises its big data potential.

Senator Deborah O'Neill

Chair