Australian Government
Australian Institute of Health and Welfare

Australia's welfare 2017

1.7 Understanding health and welfare data

Changing data landscape

In today's world of 'big data', governments, businesses, the community and individuals have access to more data than ever before. 'Big data'—the unprecedented volume, diversity and speed of data generation—is growing at a rapid pace, with the volume of digital data expected to almost triple in size between 2017 and 2020 (Productivity Commission 2017; Reimsbach-Kounatze 2015).

As its 2015 public data policy statement makes clear, the Australian Government is well aware of this trend, and of the pressing need to exploit this 'strategic national resource' (PM&C 2015a). It sees potential benefit in making non-sensitive data 'open by default' by not restricting its use or redistribution. Indeed, there may be substantial economic returns in making 'high-value' data sets (such as detailed geospatial data) more publicly accessible and enabling them to be analysed by a range of freely available tools. While the need to manage 'big data' is clearly evident, the key issue for all data users is to ensure these data are used in the most effective manner; that is, 'smart data'.

The Australian Government's agenda is to improve and, where appropriate, simplify data sharing arrangements (PM&C 2015b). It emphasises the importance of developing partnerships with government and non-government stakeholders, especially researchers (PM&C 2016a). For example, the creation of the new Data Integration Partnership of Australia in the 2017 Budget signals an ongoing commitment to '…improving policy, programs and service delivery through the better use of government data to assist in delivering a more productive economy' (DOF 2017).

There is a growing interest in unstructured data—for example, social media posts and web searches. This is being fostered by the development of machine learning techniques, where computers can track through large amounts of information for meaning, without needing explicit programming.

Privacy, security and social licence

The Australian Government's public data agenda faces substantial challenges. There is ever growing pressure to deliver high-quality, useable information faster, while ensuring that individual privacy is protected. The Productivity Commission's *Inquiry on Data Availability and Use* (Productivity Commission 2017) highlights the critical importance of obtaining 'social licence' from the community. It is suggested that, on the whole, people are willing to make their information available, but on several conditions. They need to trust how it will be handled. They need to feel that they have control over how and who will use it. They also need to see (and ideally directly benefit from) its potential value.

AIHW

Improved security is becoming more and more important as the volume of digital data expands. People who collect sensitive personal information must keep their security technology up to date. However, these steps, along with tighter privacy legislation, can only go so far to prevent breaches. These are often due to human error and cannot always be foreseen. The Productivity Commission concludes that the best approach is to assess the level of data required for different uses. In other words, risk will be minimised if sensitive information is only collected when necessary. Risk management needs to be a central focus of all data collection activity.

Increased interest in person-centred data

Alongside the mounting volume of data are growing expectations about how these data assets will be used. More emphasis is being put on ensuring that public data will be used to improve outcomes for Australian people. For example, comprehensive data are essential to develop social investment models and evaluate the impacts of programs or policies for various population groups.

Figure 1.7.1 shows the elements of policy and program areas for a person-centred data model. These form much of the 'content' of the evidence base needed to understand the experiences of the population and various cohorts within it. Across these elements, questions can be formed around:

- the characteristics of a population cohort—such as age, sex and geographical location
- determinants of the needs for support, and factors influencing outcomes—covering social, behavioural and individual factors
- interactions with the health and welfare 'system'—for example, eligibility and access, barriers (such as cost or distance), unmet need, pathways through the system
- outcomes from interactions with the system—notably education, employment and social/ participation outcomes; and health and general wellbeing
- aspects of the system that can either help or hinder intended outcomes—including policy parameters, funding models, system resources (for example, workforce levels, skills and distribution), system performance (efficiency, integration/coordination, safety/quality, responsiveness), informal carer capacity.

This model recognises that the various components may interact, and that the level of support required will differ according to individual circumstances.
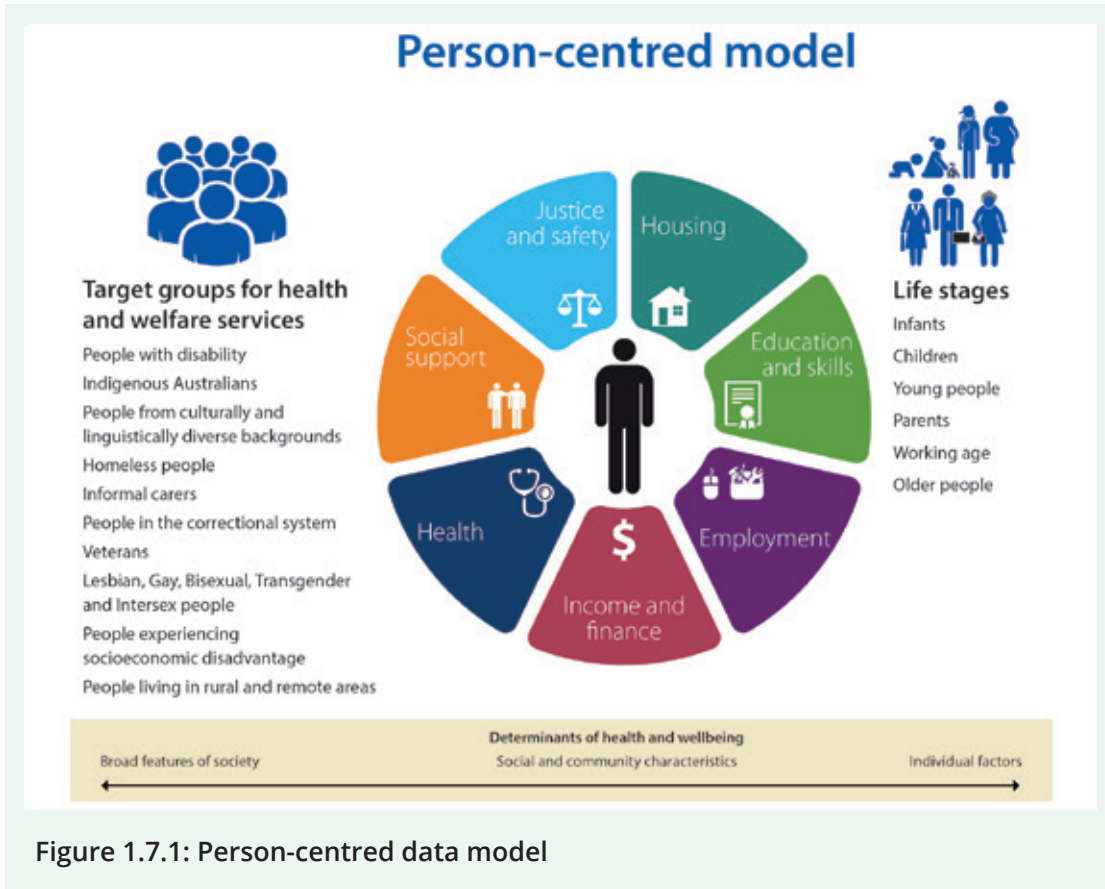
Figure 1.7.1: Person-centred data model

Types of data sources

Broadly speaking, major health and welfare data sources in Australia are either administrative data (such as collected when running a service or program) or survey data (for a targeted sample on a given topic). Each type of data has advantages and disadvantages. Choosing which one to use often comes down to purpose and the capability of each collection to measure what is needed.

The rapid growth of digitally accessible data is likely to allow much broader uses of administrative data sources (as well as to explore the potential of unstructured data). But joining administrative and survey data can provide deeper insights. This technique is used effectively, for instance, in the Business Longitudinal Analytical Data Environment. This data set combines administrative data from the Australian Tax Office with business survey data from the Australian Bureau of Statistics to tell a rich story about the characteristics and outcomes of businesses across Australia (Kalisch 2016).

Data can also be described as cross-sectional or longitudinal. A cross-sectional data source represents a particular population at a specific time. A longitudinal data source collects data on the same subjects repeatedly over time. Most Australian data collections are cross-sectional. But longitudinal studies are becoming more prevalent, as their usefulness in many policy contexts becomes more recognised. They can help governments to understand

3

how individuals respond to different situations over time; they can also identify individual pathways, and (in some instances) be directly linked to changes and outcomes after specific policy interventions (FaHCSIA 2013). Over the past 20 years, there have been concerted efforts in Australia to collect longitudinal data on a range of populations. Among these efforts was the formation, in 2014, of the National Centre for Longitudinal Data. This Centre promotes further use of longitudinal data in Australia and supports collaborative projects between researchers and policy makers (DSS 2017).

Getting the most out of data

The level and accessibility of data in 2017 is substantial. It needs, however, to be meaningfully used, not only to create evidence that informs decisions, but also to improve outcomes that can be interpreted in a policy context (as above, it needs to be 'smart' data). Data users need the skills to handle and analyse data in general; they also need to be aware of the relevant policies and contextual background (AIHW 2016a). It is essential that work in this regard is framed by strong data governance arrangements that meet legislative requirements and align with community expectations around privacy, confidentiality and data security.

Data standards

Data standards play a critical role in the meaningful use of data. 'Metadata' (data about data) allows users to have a consistent understanding of the meaning and representation of underlying data. It is a key part of making data sources as clear and usable as possible. In fact, a generally accepted principle in statistical collections is that quality metadata leads to better data.

Metadata supports consistent and transparent collection of data across national, state and territory boundaries and, in some cases, across a substantial number of agencies' data systems. For example, the Specialist Homelessness Services Collection draws on data from around 1,500 disparate non-government agencies across Australia. The comparable collation of this information supports the evidence base about people seeking homelessness services.

Data access and data sharing

The Australian Government's public data agenda is still in its early stages. But it offers many opportunities to enhance data access, paying close attention to privacy and data security. It is expected that by making non-sensitive data 'open by default', and creating integrated data sets that are widely accessible, many data gaps can be filled. Having much richer and integrated longitudinal data in future will help to answer more complex research questions. It will also better enable person-centred data to be collected and analysed. Important strategies will be to continue to improve and maximise the use of existing data sets while being proactive in identifying those that could be integrated.

All levels of government see more open data as a priority. For example, data.gov.au has been created as a central access point for a range of public data sets drawn from Australian, state/territory and local government data sets. The growing demand for individuals to control their own data is being acknowledged. This includes knowing which individuals or agencies can access this information (as seen in models such as My Health Record).

Arrangements for data sharing are progressing. One of the first major Australian Government initiatives is the Multi-Agency Data Integration Project. This involves creating an enduring, linked data set that draws information from several government data sources to answer key policy questions that can only be done by linking multiple sources. It is expected that this integrated data set will be made available to researchers via a 'trusted user' model—which allows broad data access while keeping strong privacy provisions (ABS 2016; PM&C 2016b). Some Australian and state/territory governments—Western Australia and New South Wales in particular—have already set up data integration models to answer complex policy questions, setting good examples of how multi-agency data integration models can work (Productivity Commission 2017).

A demonstration project is underway to test the linkage of a data set based on hospitals data routinely provided to the AIHW by jurisdictions with Medical Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS) data. The linked data set would be a source of patient-centred information about services provided by Australia's hospitals, medical and other services subsidised through the MBS and pharmaceuticals dispensed under the PBS. This would enhance the ability to examine patient journeys through the Australian health system.

Data linkage

Data linkage (also called data matching, data integration or record matching) is a process that allows users to combine information from multiple databases, while preserving privacy, to tell a much more powerful story than would be possible from a single source (AIHW 2014). As indicated earlier in this article, linking multiple data sources can enable more meaningful person-centred analyses, which cannot be carried out on individual data sets in isolation. Integrating data in this way can help policy makers to improve their understanding of a range of issues. These include patient- or person-centred outcomes, individuals' life courses, and the patterns of a person's interactions with various service sectors. There are many recent examples that show the substantial value that can be gained from linking multiple data sources (see Box 1.7.1 for a selection).

The number and variety of data linkage activities have increased rapidly in recent years. Three Commonwealth Integrating Authorities were set up between 2012–2014 (the AIHW, the Australian Bureau of Statistics and the Australian Institute of Family Studies). These authorities provide a secure environment in which the development of linkage projects can be fostered. Western Australia undertook over 800 projects between 1995–2016 under its WA Data Linkage System (WADLS), while the SA–NT DataLink consortium has delivered results for over 40 projects since it was set up in 2009 (DLWA 2016a, 2016b; University of South Australia 2017). Data linkage is a powerful tool. Yet challenges remain, before its benefit can be maximised. The time taken to gain approvals to use data for linkage is one pressing issue. It can affect the ability of researchers to analyse data and release results in a timely way. Another issue is the re-use of linked data for follow-up projects, a matter that is being widely discussed. Currently, most data linkage projects are funded as one-off activities. They therefore need full re-approval to use data before researchers can carry out further investigations (AIHW 2016c; Productivity Commission 2017).

Box 1.7.1: Selected recent examples of major data linkage projects

Specialist Homelessness Services and Youth Justice clients
An estimated 187,500 young people aged under 18 accessed homelessness services over a 4 year period, or about 30% of all clients. This data source was linked with the Juvenile Justice National Minimum Data Set. From this linkage, it was learned that 5,133 of these young people also had some contact with youth justice supervision. This group, compared with other specialist homeless clients of the same age, had 5 times the rate of drug and alcohol issues, double the rate of mental health issues and an overall need for more intensive support. Governments and service providers can use this analysis to design and provide more targeted responses at an individual level. For example, ensuring that mental health services are available for people exiting youth justice services and seeking homelessness support.

Human papilloma virus (HPV) vaccination program and cervical abnormalities
This study, based on Victorian HPV vaccination and Pap test data, was a world first. It showed that a population-based HPV vaccination program has produced a fall in cervical abnormalities within 5 years of its start.

MBS and the Australian Cancer Database
MBS data on 680,000 computerised tomography (CT) scans were linked to the Australian Cancer Database. This showed that exposure to CT scans in childhood increased the incidence of cancer.

Diabetes care outcomes
This pilot study involved linkage across a range of health data sources to evaluate whether new models of care deliver better quality outcomes for people with diabetes than existing practices. The study enabled several evidence-based recommendations to be developed around the integration of diabetes care and funding mechanisms.

Sources: AIHW 2014, 2016, 2016d; McKinsey and Company 2011.

Major gaps in health and welfare data

Filling data gaps

In the context of health and welfare data gaps, several themes have been identified in recent years. These include gaps in:

- the availability of prevalence data (for example, users of primary health care, and Australians who experience child abuse and neglect)
- the ability to measure meaningful outcomes for people who receive health and welfare services
- the ability to measure and track unmet demand for services
- the availability of data to measure pathways and transitions within and across different service types and across jurisdictions (AIHW 2013, 2015).

Statistical agencies work closely with data users to give priority to filling data gaps across a range of data sources. The broader accessibility of data gives governments the chance to engage with stakeholders more widely and meaningfully than ever before to better ascertain and meet their needs. For example, gaps in information collected at the local level can be determined (along with strategies to fill them) by communicating directly with Primary Health Networks and sharing relevant data with them. This strategy will allow researchers and policy experts to better target groups of interest. It will also enable them to adapt their approaches to produce much more meaningful, outcomes-based information.

Data linkage has enormous potential, but much more needs to be done to fully exploit its benefits. Data gaps in relation to services and outcomes for people with disability provide one example of these potential opportunities, and the challenges faced (Box 1.7.2).

Box 1.7.2: Filling data gaps in the disability sector

People with disability may access a range of specialist disability and mainstream health and welfare services. The creation of the National Disability Insurance Scheme (NDIS) and its subsequent national rollout highlights the opportunity and challenges of producing person centred data about services used by people with disability.

Currently, disability services data are collected at a jurisdictional level. They are collated from state/territory and Australian Government data sources via the National Disability Services National Minimum Data Set (DS NMDS). This data set provides an annual breakdown of people who access specialist disability services funded by Australian and state/territory governments. Information collected includes the characteristics and care needs of people with disability, the type and nature of support provided to them, and the mix of services they have received over the year.

The creation of the NDIS provides a chance to capture more comprehensive and meaningful data about people with disability (including outcomes) as they shift to the new model of service delivery. During the transition period, National Disability Agreement service users will continue to be captured in the DS NMDS.

The challenge from a national data perspective will be, firstly, how to capture important data about people who are ineligible for the NDIS once the rollout is complete. Then it will be a question of how best to fill long-standing gaps in available data on the use of mainstream health and welfare services by people with disability.

Future opportunities

The rapid growth in digital data will continue to accelerate. How to make optimal use of this huge and increasingly unstructured source of information is a major challenge. But it also presents an opportunity to develop better ways to analyse and present data, and to partner with a wider range of collaborators. These are issues yet to be fully dealt with. It will be an iterative process, as the full extent of how data will be used in the coming decades cannot be envisaged. Ensuring that data are used as effectively as possible will require a mix of leadership, trust and openness at all levels of government and beyond.

Where do I go for more information?

Latest news and resources on Australia's public data agenda can be found on the Department of the Prime Minister and Cabinet's Public Data web page. Currently available public data sets are also accessible via http://data.gov.au.

More information on Australia's key longitudinal data sets is available via the Department of Social Services' National Centre for Longitudinal Data.

Further information on the AIHW's data linkage program, including information for prospective researchers, can be found on the AIHW website data linking page, or by contacting the AIHW Data Integration Services Centre (linkage@aihw.gov.au).

References

ABS (Australian Bureau of Statistics) 2016. Information paper: transforming statistics for the future. ABS cat. no. 1015.0. Canberra: ABS.

AIHW (Australian Institute of Health and Welfare) 2013. Australia's welfare 2013. Australia's welfare series no. 11. Cat. no. AUS 174. Canberra: AIHW.

AIHW 2014. Data linkage services for clients. Canberra: AIHW. Viewed 20 February 2017, <http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129549730>.

AIHW 2015. Australia's welfare 2015. Australia's welfare series no. 12. Cat. no. AUS 189. Canberra: AIHW.

AIHW 2016a. AIHW response to the Productivity Commission's draft report of the Inquiry into Data Availability and Use. Canberra: AIHW. Viewed 20 February 2017, <http://www.pc.gov.au/__data/assets/pdf_file/0017/211418/subdr299-data-access.pdf>.

AIHW 2016b. Vulnerable young people: interactions across homelessness, youth justice and child protection—1 July 2011 to 30 June 2015. Cat. no. HOU 279.Canberra: AIHW.

AIHW 2016c. Australian Institute of Health and Welfare Corporate Plan 2016–17 to 2019–20. Cat. no. AUS 205. Canberra: AIHW. Viewed 20 February 2017, <http://www.aihw.gov.au/publication-detail/?id=60129556354>.

DLWA (Data Linkage Western Australia) 2016a. About the Data Linkage Branch. Perth: DLWA. Viewed 15 March 2017, < www.datalinkage-wa.org.au/about-data-linkage-branch>.

DLWA 2016b. Data linkage WA projects. Perth: DLWA. Viewed 15 March 2017, <www.datalinkage-wa.org.au/projects>.

DOF (Department of Finance) 2017. Agency resourcing Budget Paper No. 4, 2017–18. Canberra: DOF.

DSS (Department of Social Services) 2017. About the department: National Centre for Longitudinal Data. Canberra: DSS. Viewed 27 February 2017, <www.dss.gov.au/about-the-department/national-centre-for-longitudinal-data>.

FaHCSIA (Department of Families, Housing, Community Services and Indigenous Affairs) 2013. Guide to Australian longitudinal studies. Canberra: FaHCSIA.

Kalisch DW 2016. 'Big data', official statistics and national statistics offices. Viewed 15 May 2017, <http://www.abs.gov.au/websitedbs/d3310114.nsf/home/australian+statistician+-+speeches+-+big+data+official+statistics+and+national+statistics+offices>.

McKinsey and Company 2011. Evaluation report of the Diabetes Care Project. Canberra: Department of Health.

PM&C (Department of the Prime Minister and Cabinet) 2015a. Australian Government Public Data Policy Statement, 7 December 2015. Canberra: DPM&C.

PM&C 2015b. Public sector data management: July 2015. Canberra: DPM&C.

PM&C 2016a. Guidance on data sharing for Australian Government entities, March 2016. Canberra: DPM&C.

PM&C 2016b. Public sector data management: implementation report, July 2016. Canberra: DPM&C.

Productivity Commission 2017. Data availability and use, final report. Canberra: Productivity Commission.

Reimsbach–Kounatze, C. (2015). The proliferation of "big data" and implications for official statistics and statistical agencies: A preliminary analysis. OECD Digital Economy Papers, no. 245. Paris: OECD. <http://dx.doi.org/10.1787/5js7t9wqzvg8–en>.

University of South Australia 2017. List of SA NT datalink projects by financial year. Adelaide: University of South Australia. Viewed 15 March 2017, <www.santdatalink.org.au/files/PDF_Files/List_of_completed_projects20170314.pdf>.