



Australian Government
Productivity Commission

Data Availability and Use

Productivity Commission
Draft Report
Overview & Draft Recommendations

October 2016

This is a draft report prepared for further public consultation and input. The Commission will finalise its report after these processes have taken place.

© Commonwealth of Australia 2016



Except for the Commonwealth Coat of Arms and content supplied by third parties, this copyright work is licensed under a Creative Commons Attribution 3.0 Australia licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/3.0/au>. In essence, you are free to copy, communicate and adapt the work, as long as you attribute the work to the Productivity Commission (but not in any way that suggests the Commission endorses you or your use) and abide by the other licence terms.

Use of the Commonwealth Coat of Arms

For terms of use of the Coat of Arms visit the 'It's an Honour' website: <http://www.itsanhonour.gov.au>

Third party copyright

Wherever a third party holds copyright in this material, the copyright remains with that party. Their permission may be required to use the material, please contact them directly.

Attribution

This work should be attributed as follows, *Source: Productivity Commission, Data Availability and Use, Draft Report.*

If you have adapted, modified or transformed this work in anyway, please use the following, *Source: based on Productivity Commission data, Data Availability and Use, Draft Report.*

An appropriate reference for this publication is:

Productivity Commission 2016, *Data Availability and Use*, Draft Report, Canberra.

Publications enquiries

Media and Publications, phone: (03) 9653 2244 or email: maps@pc.gov.au

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au).

Opportunity for further comment

You are invited to examine this draft inquiry report and comment on it by written submission to the Productivity Commission, preferably in electronic format, by **12 December 2016** and/or by attending a public hearing.

The final report will be prepared after further submissions have been received and public hearings have been held and will be forward to the Australian Government by the 21 March 2017.

Public hearing dates and venues

Location	Date	Venue
Melbourne	21 November 2016	Rattigan Rooms Level 12, 530 Collins Street
Sydney	28 November 2016	SMC Conference & Function Centre 66 Goulburn Street

Commissioners

For the purposes of this inquiry and draft report, in accordance with section 40 of the *Productivity Commission Act 1998* the powers of the Productivity Commission have been exercised by:

Peter Harris

Presiding Commissioner

Melinda Cilento

Commissioner

Disclosure of interests

The *Productivity Commission Act 1998* specifies that where Commissioners have or acquire interests, pecuniary or otherwise, that could conflict with the proper performance of their functions during an inquiry they must disclose the interests.

Ms Cilento has advised the Commission that she is a director of Australian Unity (which made a submission to this Inquiry) and of Woodside Petroleum (which is referred to in an included example in the Report).

Terms of reference

I, Scott Morrison, Treasurer, pursuant to Parts 2 and 3 of the *Productivity Commission Act 1998*, hereby request that the Productivity Commission undertake an inquiry into the benefits and costs of options for increasing availability of and improving the use of public and private sector data by individuals and organisations.

Background

The 2014 Financial System Inquiry (the Murray Inquiry) recommended that the Government task the Commission to review the benefits and costs of increasing the availability and improving the use of data. The 2015 Harper Review of Competition Policy recommended that the Government consider ways to improve individuals' ability to access their own data to inform consumer choices. The Government has agreed to pursue these two recommendations.

The Australian Government seeks to consider policies to increase availability and use of data to boost innovation and competition in Australia and the relative benefits and costs of each option.

Effective use of data is increasingly integral to the efficient functioning of the economy. Improved availability of reliable data, combined with the tools to use it, is creating new economic opportunities. Increasing availability of data can facilitate development of new products and services, enhance consumer and business outcomes, better inform decision making and policy development, and facilitate greater efficiency and innovation in the economy.

As in Australia, international governments are encouraging greater use of data through open data policies. This will increase the transparency and accountability of government processes.

Increased sharing of data across the public and private sectors could facilitate greater leveraging of technology to improve individuals' and entities' interactions with government, improve the integrity of systems and increase administrative efficiency.

In taking advantage of greater use of data, it is important to give appropriate attention to other interests such as privacy, security and intellectual property.

Scope of the inquiry

The Commission is to conduct a broad ranging investigation into the benefits and costs of options for improving availability and use of data. In developing recommendations, the Commission is to:

1. Examine the benefits and costs of options for increasing availability of public sector data to other public sector agencies (including between the different levels of government), the private sector, research sector, academics and the community. Where there are clear benefits, recommend ways to increase and improve data linking and availability. The Commission should:
 - (a) identify the characteristics and provide examples of public sector datasets that would provide high-value to the public sector, research sector, academics and the community to assist public sector agencies to identify their most valuable data
 - (b) examine legislation or other impediments that may unnecessarily restrict the availability and linking of data, including where the costs are substantial, and consider options to reduce or remove those impediments.
2. Examine the benefits and costs of options for increasing availability of private sector data for other private sector firms, the public sector, the research sector, academics and the community. Where there are clear benefits, consider ways to increase and improve availability. The Commission should:
 - (a) identify the characteristics and provide examples of private sector datasets that would provide high value to the private sector, public sector, the research sector, academics and the community in developing or providing products and services and undertaking research and developing policy
 - (b) identify the concerns of private sector data owners and provide recommendations on principles or protocols to manage these concerns
 - (c) examine legislation or other impediments that unnecessarily restrict the availability of data, including where the costs are substantial, and consider options to reduce or remove those impediments
 - (d) provide an update on existing data sharing initiatives in Australia, including the uptake of the credit reporting framework. Consider recommendations for improving participation in such initiatives.
3. Identify options to improve individuals' access to public and private sector data about themselves and examine the benefits and costs of those options. The Commission should:
 - (a) examine how individuals can currently access their data, including data about them held by multiple government agencies, and develop recommendations to streamline access
 - (b) identify datasets, including datasets of aggregated data on consumer outcomes at the product or provider level, that would provide high value to consumers in

-
- making informed decisions and any impediments to their use. Develop guidance to assist in identification of other high value datasets
- (c) examine the possible role of third party intermediaries to assist consumers in making use of their data.
4. Examine the options for, and benefits and costs of, standardising the collection, sharing and release of public and private sector data.
5. Examine ways to enhance and maintain individuals' and businesses' confidence and trust in the way data are used. Having regard to current legislation and practice, advise on the need for further protocols to facilitate disclosure and use of data about individuals and businesses while protecting privacy and commercial interests and, if recommended, advise on what these should be. The Commission should:
- (a) balance the benefits of greater disclosure and use of data with protecting the privacy of the individual and providing sufficient control to individuals as to who has their information and how it can be used
- (b) benchmark Australia's data protection laws, privacy principles and protocols against leading jurisdictions
- (c) examine whether there is adequate understanding across government about what data can be made openly available given existing legislation
- (d) consider the effectiveness and impacts of existing approaches to confidentialisation and data security in facilitating data sharing and linking while protecting privacy
- (e) consider the merits of codifying the treatment and classification of business data.

In developing its recommendations, the Commission should take into account the Government's policy to improve the availability and use of public sector data (the *Public Data Policy Statement*) as part of its *National Innovation and Science Agenda* and to improve government performance through the *Efficiency through Contestability Programme*, as well as the findings of the *Public Sector Data Management Project*.

The Commission should consider domestic and international best practice and the measures adopted internationally to encourage sharing and linking of both public and private data.

Process

The Commission is to undertake an appropriate public consultation process, inviting public submissions and releasing a draft report to the public. A final report should be provided to the Government within 12 months from the date of receipt of the reference.

Scott Morrison
Treasurer

[Received 21 March 2016]

Contents

Opportunity for further comment	iii
Terms of reference	v
Abbreviations	xii
Key points	2
Overview	3
Access denied — Australia’s lost opportunities	4
Fundamental change is needed	12
The new Framework	13
Implementation	24
Draft findings and recommendations	25

The full report is available from www.pc.gov.au

OVERVIEW

Key points

- Extraordinary growth in data generation and usability, fuelled by developments in computing power, Internet connectivity and algorithms, have enabled a kaleidoscope of new business models, products and insights to emerge. Individuals, businesses, governments and the broader community have all benefited from these changes.
- Frameworks and protections developed for data collection and access prior to sweeping digitisation now need reform. This is a global phenomenon and Australia, to its detriment, is not yet participating.
- The substantive argument in favour of making data more available is that opportunities to use it are largely unknown until the data sources themselves are better known, and until data users have been able to undertake discovery of data.
- Lack of trust and numerous barriers to sharing and releasing data are stymieing the use and value of Australia's data.
- Marginal changes to existing structures and legislation will not suffice. The Commission is proposing reforms to data availability and use, aimed at moving from a system based on risk aversion and avoidance, to one based on transparency and confidence in data processes.
- At the centre of proposed reforms is the introduction of a new *Data Sharing and Release Act*, a new National Data Custodian, and a suite of sectoral Accredited Release Authorities that will enable streamlined access to curated datasets.
- A key element of the recommended reforms is to provide greater control for individuals over data that is collected on them by defining a new Comprehensive Right for consumers. This right would mean consumers:
 - retain the power to view information held on them, request edits or corrections, and be advised of disclosure to third parties;
 - have improved rights to opt out of collection in some circumstances; and
 - have a new right to a machine-readable copy of data, provided either to them or to a nominated third party, such as a new service provider.
- Broad access to key National Interest Datasets should be enabled.
 - For datasets designated as national interest, all restrictions to access and use contained in a variety of national and state legislation, and other program-specific policies, would be replaced by new arrangements under the Data Sharing and Release Act.
 - Datasets would be maintained as national assets, access would be substantially streamlined, and linkage with other National Interest Datasets would be feasible.
 - Initial datasets that may be designated national interest and publicly released could include key registries of businesses, services or assets, and data on activity and usage in areas of substantial public expenditure.
- Secure sharing of identifiable data held in the public sector and by publicly funded research bodies should be formalised and streamlined. By pre-approving data uses, trusted users would have more timely access to identifiable data through Accredited Release Authorities and ethics committees.
- Reforming access to public sector data is a priority. Significant change is needed for Australia's open government agenda to catch up with achievements in competing economies.
- The incremental costs associated with more open data access and use — including possible impacts on individuals' privacy and willingness to share data — are expected to be minimal, but they will exist. But greater use of Australia's data can coexist with the management of these risks, including genuine safeguards and meaningful transparency to maintain community trust and confidence.

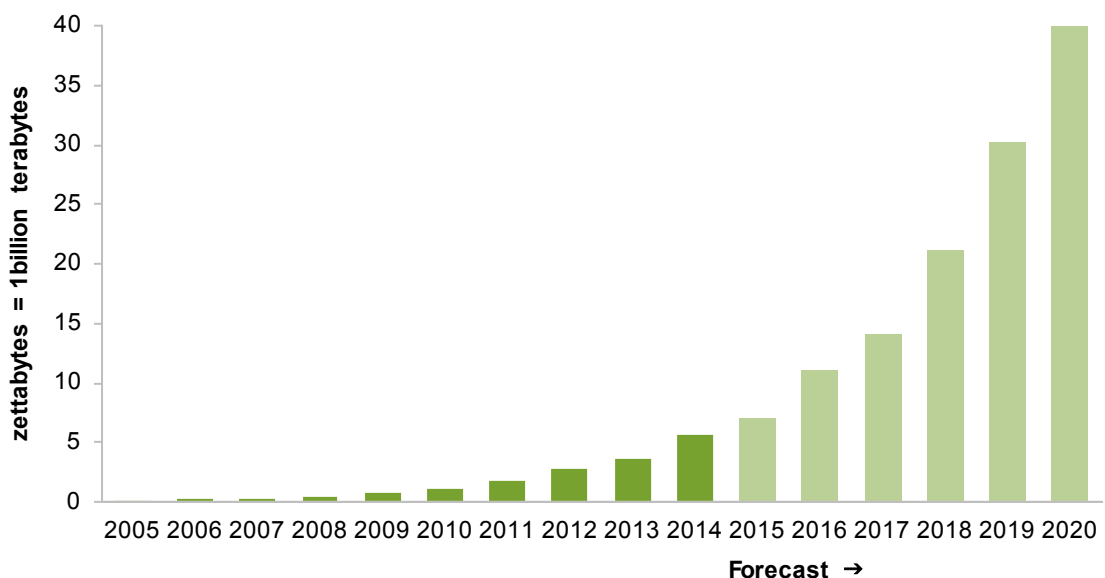
Overview

Thirty years ago, data for most people was primarily about details on paper. Data was largely collected and stored on paper (encyclopaedias, forms, bills, bank deposit slips and phone books); mail actually meant a letter in the letter box. Access to data was clear and locational (you needed keys to the filing cabinet); as was its destruction (via a shredding machine). With the mass digitisation of data, the capacity to collect data through everyday Internet activity and transactions, and through technologies such as sensors, cameras and mobile devices, means that what is ‘data’, and who can or should have a say in how it is collected, stored and used is no longer so simple.

Until this Inquiry, there has been no structured attempt to comprehensively review this matter in Australia, despite the enormity of the transformation under way.

Data now includes material (raw or processed) on: the characteristics, status, appearance or performance of an individual, product or service, or object (including infrastructure and environmental assets); and expressed or inferred opinions and preferences. The generation of data is seemingly heading upward on an unbounded trajectory (figure 1).

Figure 1 **Data generated (global)**



Source: United Nations Economic Commission for Europe (2015).

By some estimates, the amount of digital data generated globally in 2002 (five terabytes) is now generated every two days, with 90% of the world's information generated in just the past two years (IBM 2016). As we are now only in the very nascent stage of the Internet of Things (whereby our business equipment, vehicles, appliances and wearable devices can communicate with each other and generate data), the upward trend in data generated is more likely than not to accelerate into the future.

Falling costs (per record) of digital data storage, and the spread of low-cost and powerful analytics tools and techniques to extract patterns, correlations and interactions from within data, are also making data analytics more usable and valuable. Yet much of the data being generated remains underutilised. Some estimate that up to 80% of data generated globally may prove to have no value (numerous duplicative digital photos, for example). But still, less than 5% of the potentially useful data is actually analysed to generate information, build knowledge, and thus inform decision making and action (EMC Corporation 2014). And some data that was previously of limited value is becoming valuable as new uses for it emerge, analytical capabilities improve, new linkages are established, or investments made to improve its quality. There is enormous untapped potential in Australia's data.

Access denied — Australia's lost opportunities

With technological developments and advances in analytical techniques, not only is the volume of data being generated and collected growing, but so too is the scope to make use of data in innovative ways in every sphere of life.

Increased access to data can facilitate the development of ground-breaking new products and services that fundamentally transform everyday life. Many are widely known — apps that tell you in real time where to find vacant car parking places, the fastest route to travel to the city at the time you want to go, or which electricity provider offers you the best deal given your pattern of energy use, are all examples that rely on data analysis.

But better access to and use of data can also benefit business and government through improved operational processes and productivity. Examples abound of new found opportunities — in supply chain logistics, saving time and money; through more cost effective infrastructure and machinery maintenance and planning; improved safety and efficiency in aircraft engines; and in the capacity to better respond to and manage emergencies. And data is critical to building the evidence base to underpin incremental improvements, allowing governments and businesses to offer products and services that are more customised, coordinated or timely. The potential value of data is tremendous, but so too is the scope for Australia to forgo much of this value under the misconception that denial of access would minimise risks.

While this Report highlights some examples of where data is already being used to benefit the community, these are the tip of the iceberg of what could be achieved. What is already being done with data overseas is indicative of what is possible in Australia, if only more data could be released for use and the risks managed.

Health data exemplifies the problem

Australia's health sector exemplifies many of these opportunities, to date largely foregone, due to impediments and distrust around data use (box 1). Data from the sector that could be more widely used includes:

- broad level performance data on expenditure and activity at particular medical facilities (the number of available public and private hospital beds by state and territory) for particular medical conditions (the number of people diagnosed with asthma in each of the past 20 years and public expenditure on particular types of asthma treatment);
- finer level performance data on particular parts of the sector (the number of serious complications following orthopaedic surgery at each hospital, or how drugs prescribed for particular medical conditions vary across medical practitioners);
- data that relates to the health records of individual patients (documented reasons for visits to health professionals, the results from diagnostic testing undertaken, prescriptions received, private and public health insurance claimed); and
- data collected through personally controlled devices, such as smartphones and health monitors, that have an increasing potential to assist medical practitioners and patients themselves.

From the Commission's experience with its annual *Report on Government Services*, data that allows performance monitoring and comparison of government activities is a fundamental starting point for improving the delivery of those activities to the community. While data in that publication motivates a closer examination of practices within particular sectors and jurisdictions, the highly aggregated level limits its use by governments, businesses and the community in making better informed decisions about health products and services. Yet behind many of these thousands of aggregated data points are datasets, the equivalent of which capable, trusted researchers in nations — the United States, New Zealand and the United Kingdom — can and do actively analyse to enable discovery and solution to seemingly intractable problems. And in that context, we fall short.

Inquiry participants highlighted a range of health sector data that could underpin substantial long lasting benefits for the Australian community.

Using data to anticipate and prepare for community and individual health needs

Health data can help policy makers and researchers to:

- identify emerging health issues within communities and factors that contribute to particular medical conditions;
- assess the safety of pharmaceuticals and other treatment options on an ongoing basis; and
- evaluate the effectiveness and efficiency of health policy.

Box 1**Australia's health data — an underutilised resource that could be saving lives**

Due to a multitude of legal, institutional and technical reasons, Australia stands out among other developed countries as one where health information is poorly used (OECD 2015c):

The health sector is very good at generating and storing data. It is less effective at translating this data into useful information. It is poor at linking and sharing information between health professionals, where it could be used to improve health outcomes and system efficiency. Worst of all is the health sector's ability and willingness to share data with consumers (Medibank Private, sub. 98. p. 2).

The implications of this situation are significant. At the individual level, patients are required in many cases to act as information conduits between the various health care providers they see. Inadequate information can lead to errors in treating patients (Joint Council of Social Service Network, sub. 170). At the system level, inefficient collection and sharing leads to data gaps and unnecessary expenditure:

- [H]ealthcare providers largely operate in disconnected silos, hindering continuity of care. Doctors often do not know what medications and tests have been given to patients by other doctors, even when they are members of the same care team. It is even more difficult to bring relevant medical knowledge to the point of care, to create integrated care plans, to monitor a patient's progress against the care plan, or to alert care providers when a patient's condition requires intervention. (Georgeff 2007, pp. 6–7)
- A Parliamentary Committee in Western Australia reporting on data portability problems at one hospital stated “the Health Services Union indicated that the ICU CIS was not compatible with the systems in use on the general wards. According to the HSU, this meant that patient's records must be printed and scanned when they transfer from the ICU to a general ward.” (Education and Health Standing Committee (Western Australia) 2015, p. 23)

Furthermore, the lengthy approval process for researchers requesting access to personal data limits their ability to make potentially life-saving discoveries:

- Nearly five years after requesting the data, researchers at the University of Melbourne received de-identified information about CT scans and cancer notifications. Their work showed there was an increased cancer risk for young people undergoing CT scans, and led to changes in medical guidelines for the use of scans. “Had [the] study been approved sooner, and been able to proceed at an earlier date..., we would have had results sooner, with potential benefits in terms of improved guidelines for CT usage, lesser exposures and fewer cancers” (John D Mathews, sub. 36, p. 13).
- Since 2008, the Australian Research Council and other government bodies have been providing funding to the Vaccine Assessment Using Linked Data Safety Study. Among other objectives, this study examines whether there is a relationship between vaccination and admission to hospital or death. The study requires data from both the Australian and State Governments. Obtaining data from the Australian Government has taken six and a half years; state data has not yet been linked. According to Research Australia (sub. 117), linkage is expected to occur in late 2016, eight years after the project commenced.

Electronic health records, for example, could incorporate and use data from monitoring devices to help to identify patients most likely to benefit from particular interventions, and predict those patients whose condition is likely to worsen (which would allow for targeted interventions by healthcare providers).

In the UK, administrative hospital records linked (via unique patient health service number) with a number of cancer screening registries have been used to improve how and when cancer is diagnosed (to increase early detection and survival). Undertaking similar analysis in Australia would require linking of data held by a range of groups, including data from Medicare Australia, the Australian Government Department of Health and its counterparts in the states and territories, various cancer registries and other organisations.

There is already strong support for using Australia's health data in research. A recent survey revealed that over 90% of Australians were willing to share their de-identified health data to advance medical research and improve patient care (Research Australia 2016). Yet more effective use of data is not being sufficiently enabled. Inquiry participants noted a wide range of further medical advances and health sector transformations that could be made possible through the linkage of administrative data with large scale health data collections (such as the 'Busselton Health Study', the 'Australian Atlas of Healthcare Variation', and '45 and Up'), and private sector health insurance data.

Data that allows improved service provision

Inquiry participants flagged the potential for data relating to health service provider costs and performance, as well as de-identified linked data about health service recipients, to be used for more effective and targeted service interventions and improved health outcomes.

The New Zealand Treasury has used longitudinal data from anonymised linked administrative datasets (in this case, mental health program usage and pharmaceuticals) to identify young people who are at risk of poor outcomes in adulthood. By identifying a number of key characteristics that appear predictive of poor future outcomes, the analysis provided valuable insights into the effectiveness of various policies and interventions. The separation of data holdings across three levels of government and across different agencies within each of these jurisdictions, and the distrust that inhibits sharing of this data for linkage purposes, means that such analysis is not yet feasible in Australia.

Yet opportunities are emerging. Integrating data across clinical systems is becoming feasible with greater adoption of electronic health records in Australia. This would enable more effective and holistic healthcare for patients who receive treatment from a range of healthcare providers. While some duplication of diagnostic processes may be necessary for certainty or for alternative treatment plans, roughly 10% of pathology and other tests have been found to be unnecessary duplicates (CBO 2008). As using data to alert practitioners to duplicate radiology tests has been estimated to reduce the number of tests by up to 25% and test waiting time by up to 50% (Chaudhry et al. 2006), there are substantial gains in service efficiency to be had from reducing duplicative effort and integrating health data.

To allow new services to emerge in response to community demand and compete with existing product offerings, potential providers need geographic information on current use of health services. The Australian Dental Association highlighted that access to private health insurance data could allow for new dental practices to be established in areas of high demand.

Data that empowers individuals in managing their use of health services

Patient access to their own medical history (wherever they are, instantly) would not only improve professionals' knowledge of their patients' medical condition and reduce the number of diagnostic tests, but enable the ready and secure sharing of health information to other healthcare providers.

Some private sector services are already developing in Australia to allow consumers to manage their health data. Health&, for example, allows consumers to manually input and store their health data, including medical records and data from fitness devices, in a centralised location to allow better preventative health care and simpler sharing of health information between health service providers. How much more efficient and less error-prone would such transfer be if this could be done at a key-stroke? And it can, but not in Australia. That such services exist, even though they rely on manual rather than electronic input of information, is indicative of the appetite of some consumers for more control over the management of their own health data.

The risks are real but manageable

Allowing and enabling data to be available and used more widely would provide enormous benefits, but there are risks involved. These risks vary with the nature of the data holding, and the environment and purpose for which it is used. Release of aggregated data on government regulatory activities, for example, may pose a very low risk of adverse consequences. Release of data that identifies individuals who have attended a particular medical facility could, in contrast, be highly detrimental to both the individuals concerned and the reputation of the facility. Thus, the risk of harm needs to be assessed based on both the likelihood and scale of harm associated with data being more widely available. Where the adverse consequence of increased data access are considered high, the availability of the data needs to be carefully managed.

The types of risks that Inquiry participants pointed out as being most significant — related to the potential to identify persons or businesses within datasets — were:

- discrimination
- loss of control over the boundaries around the 'you' that the world sees
- reputational damage or embarrassment
- identity fraud
- other criminal misuse of the data
- commercial harm.

That these risks exist is undeniable, but it is important not to fall victim to fear. Some, indeed most, apply to every form of data management, including pen and ink. Identity theft affected 126 000 Australians in 2014-15 (ABS 2016e). Most personal information used in identity theft is obtained online, either through theft, hacking or from information sent by email or

placed on a website, rather than through data release or sharing. Some victims have suffered financial losses; others have reported being refused credit or accused of a crime.

Risks of identification can increase with the linkage of separate pieces of data about an individual. Matching data across individuals can also reveal more information about the activities and associations of those individuals.

These risks — and the desire for privacy and confidentiality — should not be downplayed or trivialised. They are real and important. But, many of them are able to be managed with the right policies and processes. The likelihood of unintended or inappropriate release needs to be carefully considered alongside the likelihood of any genuine harm or costs to the individuals or organisations concerned. Systems and processes can and should be developed to identify, assess, manage and mitigate risks related not just to data release and sharing, but also data collection and storage. Where it is not possible to reduce risks to an acceptable level, the approach being advocated by the Commission would not support release of the data.

Even with data that has never been about individual persons or businesses — such as data on the use of publicly funded facilities — increased availability can come with risks of misuse, where the quality, meaning or context of the data is not understood by users.

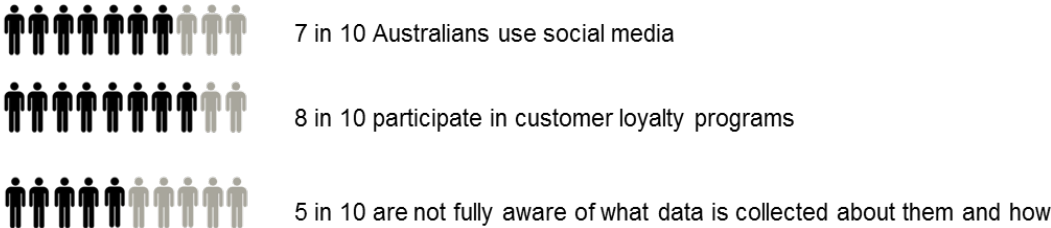
Giving data away

Australians give away a lot of personal information online (figure 2). For many, the information gate is (often consciously) wide open. In innumerable ways, individuals deliberately or inadvertently provide information about themselves for one purpose, which then is, or has the potential to be, used for other purposes.

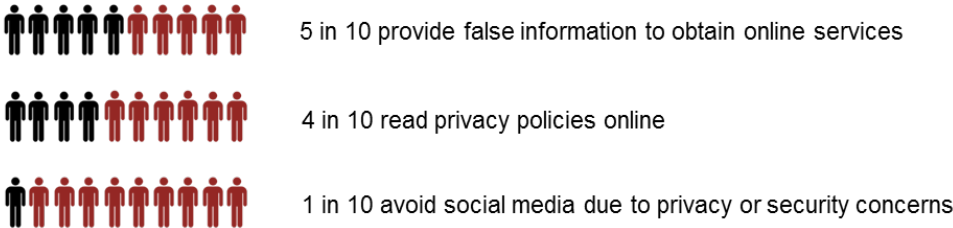
- Some 68% of Australian Internet users have a social media profile, with one quarter accessing their account more than five times per day. The most popular of these sites, Facebook, soaks up information from users' computers and uses it to earn 96% of its revenue through targeted advertising. Only 12% of Internet users avoid social media for security or privacy reasons.
- Similarly, around 84% of Australians are enrolled in at least one customer loyalty program — with an average of 3.8 program memberships. While 47% recognised that a primary reason for loyalty programs is data collection by the company, less than 2% were concerned about their privacy or felt the business knew too much about them.
- Australians have a relatively big appetite for technologies that generate or collect data (we are typically early adopters). For example, at 13% of the population, Australia has the second highest take-up rate of fitness band devices in the world. Wearable technologies, such as Fitbits, transfer data on the physical wellbeing and location of individuals back to the device provider and may be reused by it.

Some 47% of Australians report altering personal information provided online in an attempt to make themselves less identifiable (ACMA 2013a), but ignore the fact that fragments of correct information on them from a wide variety of sources are being compared and matched by intelligent algorithms to form a complete and accurate picture of them.

Figure 2 The risks that Australians take with data



Only some take action to protect their privacy



Source: Directivity et al. (2015); ACMA (2012, 2013a); OAIC (2013a); Sensis (2015)

That privacy is often said to be a concern but individuals still willingly and readily hand over personal information may seem a paradox. Because much of the data that is being generated is a byproduct of other activities, it was once easy for individuals to dismiss it as being of secondary importance. Today, that should not be the case. If you are using a product or service and not paying for it (or sometimes even when you are), then you are the product. This is perhaps most obvious by the ‘all or nothing’ nature of personal data requested in exchange for typically free online products and services. What you are consuming, how and when you are consuming it, is all being collected as data that is likely of more value to the supplier than whatever it is they are offering you.

Individuals typically have less choice about providing personal information to governments and may see a less immediate or personal benefit from doing so. Despite claims of a few privacy advocate groups, this Inquiry has not been presented with evidence to suggest widespread concern about the provision of personal information to governments. Indeed, the Office of the Australian Information Commissioner has found that 70% of Australians trust governments in the handling of personal information (only health service providers and financial institutions were rated higher). If individuals do have concerns about provision of personal information to governments, we would welcome hearing these

views for further consideration in the Inquiry final report. The Productivity Commission website is established to receive comments (<http://www.pc.gov.au/inquiries/current/data-access/make-submission#lodge>).

Increasing data use does not necessarily increase risk

In reality, most risks of data misuse arise not through the public release of robustly de-identified data, but rather from poor or outdated data collection, storage and management practices, often coupled with malicious intent to gain access and use data that would otherwise not have been available. The other avenue made possible by increased online activity is misuse of personal information that individuals have readily made public, to access other information that is not public (essentially a form of identity theft). As the value of data rises, the incentives for such exploitation rise, underscoring the need for all data collectors to remain vigilant and up-to-date in technology around data collection, handling and de-identification.

Given these sources of risk, the main factors stopping breaches of privacy are safeguards around data handling — prompted by the desire of most large private sector data holders, ethics committees, and public sector data custodians to maintain trust and reputation — and inaction on dataset release to avoid potential legal recrimination, given profound uncertainty about privacy and secrecy requirements.

Yet this inaction not only denies discovery (and perhaps innovation), it also takes no account of incentives — for example, there is a profound lack of interest amongst most researchers in government and academia in identifying particular individuals from large datasets; for them, de-identified datasets about large *groups* of people hold the answer to many pivotal questions.

That most data breaches are inconsequential and go largely unnoticed is hardly the point. It only takes one major breach to destroy public confidence. But tightening privacy legislation will not prevent human error and is, at best, a small disincentive to criminal intent.

Greater use of data does not mean Australians should be put at greater risk of harm. In fact, it is vital for Australia's data future that the risks of data handling are managed. The re-identification problems with the recent release of de-identified linked MBS-PBS data underscore the clear need for a robust framework, including expert technical support, for the management of Australia's high value public interest datasets. Against the background of an ocean of personal data that is already public, there remains a need for continued community acceptance and trust in the handling of personal data by governments and business. Built through genuine safeguards, meaningful transparency, and effective management of risk, such acceptance and trust will be vital for the implementation of any reforms. This should be the overarching objective of any reform agenda.

Fundamental change is needed

The legal and policy frameworks under which public and private sector data is collected, stored and used (or traded) in Australia are ad hoc and not contemporary. The impetus for changes in governance structures around data — changes that deal head-on with the fact that data is increasingly digital, revealing of the activities and preferences of individual people or businesses, and held in the private sector — will not diminish. It is a global movement and, to its detriment, Australia is not participating.

Tweaking existing structures and legislation will not suffice. Rather, fundamental and systematic changes are needed to the way Australian governments, business and individuals handle data. This conclusion is based on a number of findings:

- The nature of data sources and data analytical techniques are evolving rapidly and moving away from any effective control by individuals, and will continue to do so — doing nothing is no longer an option.
- As data standards and metadata improve, digital data will be able to be transferred across the economy, between sectors and across national boundaries with increasing ease. To ensure coverage is comprehensive and understandable, data management frameworks need to be consistent across the economy.
- Incremental changes in the data management framework to date have failed to deliver a culture of making data available for widespread use. The range and volume of datasets now held in the public or private sector, that *could* potentially be made more widely available and the associated opportunities are monumental. While there have been noticeable increases in the sharing and release of certain data in recent years, these releases remain a ‘pimple on the pumpkin’ of data release possibilities.
- There are key unanswered questions that go to the fundamental rights of individuals to data held about them, and how individuals can use data more effectively for their own benefit, that lie at the heart of data availability and use. These questions necessitate an across-the-board rethink of the way data is managed.

The Commission’s recommended approach incorporates recent progress in policy and practices around data management but is deliberately aimed at creating a new, comprehensive framework that should, by design, be capable of enduring beyond current policies, personnel and institutional structures. It takes account of the significant differences in data types and associated risks and uses of each, and recognises that while the incremental risks of making data more available might appear very small (given how much data is already in the public domain), incentives and trust nevertheless have to be maintained. Crucially, the proposed reforms take Australia beyond the stage of viewing data availability solely through a privacy lens, in recognition that there is much more than privacy at stake when it comes to data availability and use.

The new Framework

A key issue in balancing access and trust is consideration of the level of data required for different uses. Near real time data that identifies individual persons or businesses carries the highest risks to privacy and security. Access to this level of data by those other than the parties to a transaction — while useful for the enforcement of some regulations (for example, traffic speed limits) and for inducing timely changes in consumer behaviour (for example, price responsive household electricity consumption) — is not necessary in order to obtain much of the benefits of data use. For analysis of market opportunities, scenario development, policy evaluation or improved delivery of many products and services, de-identified data can be sufficient, and indeed, desirable. And, of course, there is considerable data that is non-personal and non-confidential, which also needs to be made more accessible for use and reuse.

The Commission's Framework (figure 3) recognises the spectrum of risk associated with different types and uses of data, and the corresponding need for different risk controls and approaches to apply.

Where the risks associated with release cannot be effectively mitigated, the Commission's approach would not involve release in the case of *genuinely* commercial in confidence data, or data that is integral to the security of the country. For the remaining bulk of data, the recommended approach to improving sharing or release is detailed below, and reflects a sliding scale of release strategies and controls commensurate with the potential risks and benefits of potential release.

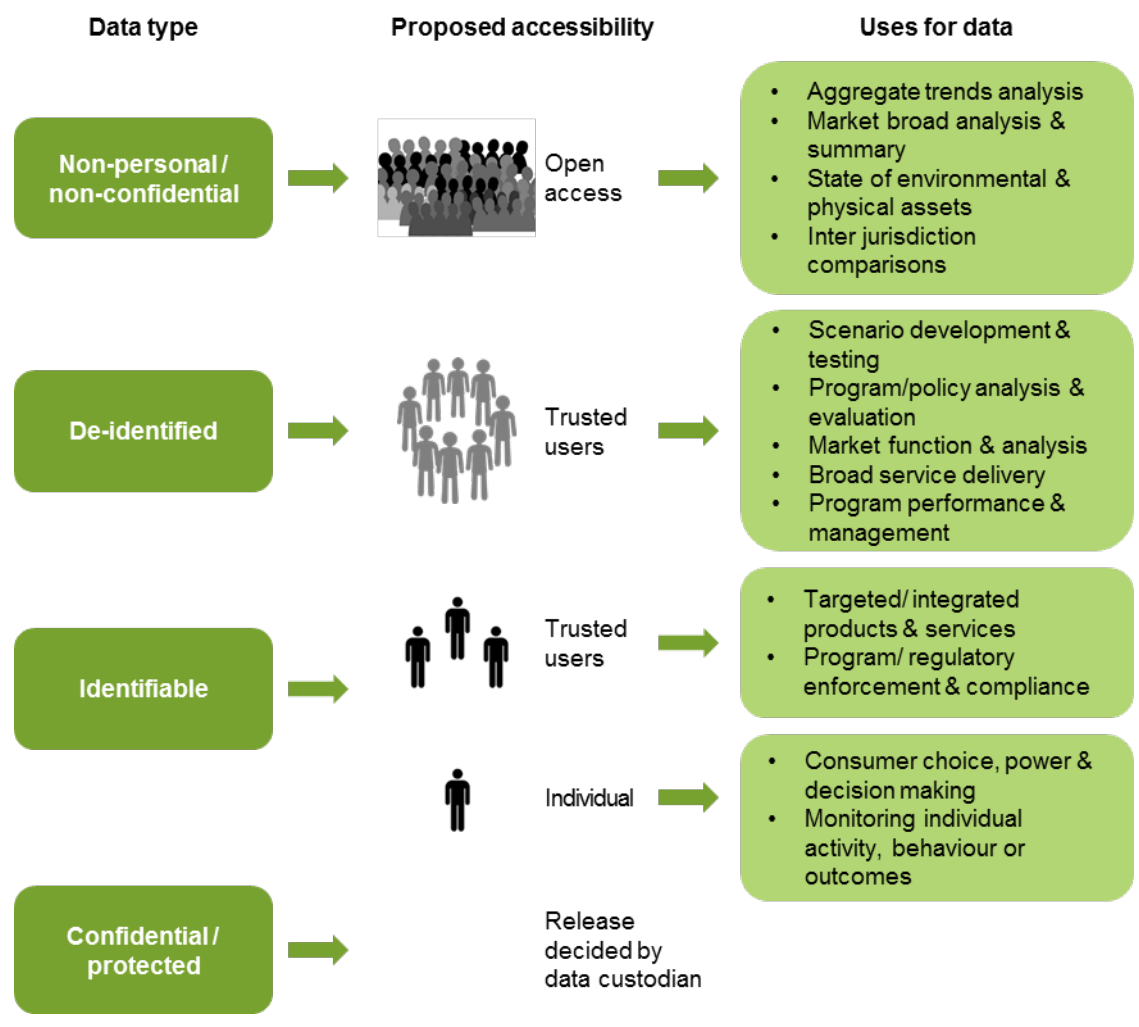
Broad criteria shaping the Framework are that it must: deliver net benefits to the community; increase the availability and usefulness of data; engender community trust and confidence in how data is managed and used; and preserve commercial incentives to collect, maintain and add value to data.

There are four key elements to the recommended approach, consistent with the Inquiry terms of reference, and underpinned by changes to legislative and governance structures:

1. Giving individuals more control over data held on them
2. Enabling broad access to datasets (public and private sector) that are of national interest
3. Increasing the usefulness of publicly funded identifiable data amongst trusted users
4. Creating a culture in which non-personal and non-confidential data gets released by default for widespread use.

Each of these elements is discussed in detail below.

Figure 3 Framework of the recommended approach



New legislation and governance structures for data access

Although this Inquiry would have preferred to find solutions that are non-regulatory, it is a clear conclusion that legislative changes are needed to implement the Commission's recommended reforms. These primarily involve changes to existing Commonwealth privacy legislation as well as the creation of new legislation — a new Data Sharing and Release Act — to facilitate data sharing and release. This Act would be a Commonwealth piece of legislation *applying across Australia to all digital data*. It would therefore be ‘umbrella’ legislation, and would make redundant some clauses in other dataset-specific and program-specific legislation around privacy, secrecy and other matters. Apart from giving entities ‘permission’ to publicly release data while managing risks, it forms the basis for a new lens through which to view data availability and use: the lens of a valuable asset being created, not merely a risk or an overhead.

Further consideration will be particularly needed in regard to the interface between the new Act and the Privacy Act. A primary intention would be to retain the key protections within the latter legislation, particularly as they apply to the use of personal information, whilst also ensuring that the new Act facilitates a more open and effective approach to data management.

Implementing this legislated Framework would be a central government agency with data responsibility, a new national statutory office holder — the National Data Custodian (NDC) — and a suite of sectoral Accredited Release Authorities (box 2). These Accredited Release Authorities (ARAs) would be funded and tasked with assisting data custodians to improve the curation and quality of datasets to be released (including de-identifying data where necessary), and facilitate timely updates and ongoing dataset maintenance.

The ARAs would ultimately, on advice from original dataset custodians, be responsible for deciding whether a dataset is available for public release or limited to sharing with ‘trusted users’. Trusted users would *potentially* include any individual working in an entity that is covered by privacy legislation, with necessary governance structures and processes to address the risks of inappropriate data use or release associated with particular datasets, including access to secure computing infrastructure. Trusted user access to higher risk identifiable data would necessarily be more tightly controlled, with more stringent requirements on the means and purpose of data use to be satisfied before the ARA could grant access.

In addition to these new structures, additional roles may be afforded in governance to the Office of the Australian Information Commissioner, the Australian Competition and Consumer Commission, the Administrative Appeals Tribunal, and relevant industry ombudsmen. Bodies such as the Australian Bureau of Statistics, Australian Institute of Health and Welfare, Sax Institute and CSIRO Data61 may also take on additional technical advice and support roles with regard to data linkage and integration, de-identification, and safe sharing and release.

Giving individuals more control over their data

Australian consumers have little capacity to choose how personal data about them is used; and too often, organisations and governments make decisions about the use of individuals’ data on behalf of the individuals concerned. In the face of the ubiquity of data collected, the scope to provide consumers with a greater say — within limits — on the handling of data that is sourced from them, is considerable.

Box 2 **Key institutions and roles in the new Framework**

The Commission's recommended reforms require government to take up important new functions to enable the opportunities from data to be realised. This will require the establishment of a new national position, and the authorisation, through the new Data Sharing and Release Act, of some additional functions by existing institutions.

The National Data Custodian (NDC)

The new position of the NDC creates a role that parallels for data access and use, the role of Australia's Information Commissioner for data protection. The NDC will have responsibility for broad oversight of the operation of the national data system, be involved in designation of datasets of national interest, potentially with designation by disallowable instrument. The NDC would also accredit release authorities and trusted users within the reformed data system.

Accredited Release Authorities (ARA)

ARAs will largely be existing public sector agencies (Australian Government or state/territory government) that already release data but would now be funded to take on additional responsibilities as an ARA (the Australian Institute of Health and Welfare may be a workable model for an ARA). ARAs would play an important role in deciding whether a dataset is available for public release or limited sharing with trusted users, curating datasets and assisting dataset custodians with curation and the development of metadata, ensuring the timely update and maintenance of datasets, and supporting the linkage of national interest and other datasets for release. Given the emphasis on sectoral expertise, these entities would have a long track record of trusted data management in their particular areas of focus. It is envisaged that ARAs would also perform an important advisory role on technical matters, both to government, and to the broader community of data custodians.

Other existing institutions with additional roles

Existing institutions with important new roles in the reformed data management system include:

- Australian Competition and Consumer Commission — oversee the Framework around consumer access to data
- Office of the Australian Information Commissioner — continue its fundamental role regarding privacy and handling of privacy related data complaints
- Administrative Appeals Tribunal — possible role to assess disputes and appeals regarding data sharing and release
- Australian Bureau of Statistics, Australian Institute of Health and Welfare and CSIRO Data61 — key advisory and support roles, in addition to their existing functions within the data infrastructure.

Control and access to personal information can help encourage information sharing because it builds individuals' confidence that their personal information will be used in a way that reflects their preferences. Increased access and greater control over their ability to use data collected about them also affords individuals more choice about the products and services they consume (and the providers of those), and is an avenue to improve market competitiveness. Currently, it is a business or service provider that determines how to extract value from an individual's data. We propose to explicitly create the opportunity for

individuals also to choose to trade or reuse their data. While consumers arguably already have some access power, there are severe practical constraints in Australia at present, on how to exercise this. But this need not be the case.

A new right to data access

Under the Commission's proposed plan, consumers would retain the existing ability to view what information a business or government agency holds on them and request edits or corrections (related to accuracy). The capacity to have data edited would be a right to *request* specific edits, not a right to compel data holders to change their datasets unless incorrect; however, it would generally be in the data holder's interests to ensure data holdings are as accurate and up-to-date as possible. This design feature is necessary to ensure that inaccurate corrections are not made: the Wikipedia experience is best not applied to health data.

Consumers would also have a right to be informed of disclosure of data by a data holder to third parties; and a right to appeal automated decisions, such as those based on statistical profiling. *The Commission's recommendation increases powers of individuals and formally defines them as a 'right', while maintaining safeguards.* This access right would be enforceable in the same way that existing powers are — via complaint to the Office of the Australian Information Commissioner (OAIC) or the relevant industry ombudsman.

Under the Commission's recommended approach, consumers would also have an explicit new right to require that a data holder stop collecting information on them (that is, they can 'opt out' of a collection process). This capacity to opt out at any time would be subject to a number of exceptions, including that individuals would not be able to have collection cease if the collected information is necessary for public benefit purposes (such as the maintenance of public health and safety, or administrative purposes such as tax collection) or forms part of a National Interest Dataset (described later).

In the private sector, opting out of data collection may well mean that a particular product or service is no longer available or no longer free to that consumer. But consumers must, in the ever-expanding world of data opportunities, be able to make that call for themselves.

Nor would the right to stop data collection include having historical data deleted or use of it cease. This provision recognises that once data is integrated into a dataset or analysis, it can be costly or infeasible to extract it, and on occasion, damaging to the interests of others using the dataset. It is also intended to ensure that individuals' opt out decisions do not decimate the investment that data holders have made in datasets and, in some cases, ensure that information on past activity is available to inform future activity (information on past medical procedures of an individual would be necessary for future medical treatment, for example).

More scope for individuals to use data about themselves

The new Comprehensive Right of individuals over their data would extend to include the ability to direct that a copy of their data be transferred safely from one data holder to another. This is a key additional power afforded to individuals under the new Framework.

The capacity for individuals, as consumers, to copy their data between service providers is an integral part of facilitating competition in markets and reducing barriers to market entry. In some circumstances, the consumer may see benefits in having a copy of their data provided to an entity that is not a competitor (for example, provision of medical records to a life insurance company or provision of utility payment information to a credit provider). In other cases, it will be to form a new customer relationship, or obtain a quote that may lead to one, at the consumer's discretion.

Underlying this right, and to maintain incentives for existing data holders and collectors, is the idea that the right to data is a *joint right*, shared between the individual and the businesses or agencies that hold the individual's data. The individual's decision to switch service providers would not alter the right of the initial service provider to retain the data that they had already collected while providing a service to the individual.

All businesses, government agencies and government business enterprises should be subject to this new Comprehensive Right. In some sectors, transfer of data may be achieved by the use of application programming interfaces; in others, the transfer of files. Either way, standards around data formats and definitions will be necessary. We consider that participants in each sector, rather than governments, are best placed to develop these standards. But we propose a process to achieve this that draws on existing standards development practice in Australia.

Any charges levied by data holders for access, editing, copying and/or transferring of data should be monitored by the Australian Competition and Consumer Commission (ACCC), with the methodology transparent and reviewable on request by the ACCC.

While the changes proposed aim to enable consumers to exercise more control over the collection and use of data on them, the onus remains on individuals to make responsible choices about whom they provide personal information to in the first instance and for what purposes.

Comprehensive credit reporting

In some circumstances, collating consumer data may offer net public benefit in making markets more efficient. A specific case is covered in the terms of reference for this Inquiry: comprehensive credit reporting. The Productivity Commission has previously found comprehensive credit reporting to be desirable and, consistent with the approach of New Zealand, the United Kingdom and the United States, a voluntary approach to data input

should continue to be pursued, unless it becomes clear that a critical mass of accounts is not achievable on that basis.

Broad access to datasets of national interest

We have given considerable thought to establishing an element in the Framework to enable wider access to high value, National Interest Datasets. The intention is to promote the development of a valuable suite of datasets — some of which are released publicly; others that will be shared with a smaller group of trusted data users. Designating datasets as national interest collections will also signify their value as resources collected in the national interest, not merely (as today) for compliance, record-keeping or audit.

The term national interest in our Framework necessarily covers data with a significant public interest element that is collected by Australian, state/territory or local government agencies and publicly funded research bodies. In specific cases, private sector data deemed to have a public interest element will also form part of a National Interest Dataset — private health insurance data is a potential example.

Governments across Australia hold enormous amounts of data, but lag behind other comparable economies by typically not exploiting it beyond the purposes for which it was initially collected, nor allowing others access to do so. Australia's private sector data holders are more innovative in their use of data. But even so, the extent to which their data holdings are available more broadly — when data are collected through public funding or to meet a public interest objective — remains constrained by limitations on data linkage, ad hoc frameworks to facilitate release and commercial incentives. Wider availability of such data (public and private sector) would likely trigger significant investment and improvements in national welfare.

The extent to which use of a particular dataset could provide benefits to the broader community (beyond those derived by just the data holders and data contributors), is important in considering how widely available it should be made — that is, whether the objective is ultimately to release the data, or to share it more effectively.

While there are some important and obvious initial examples of likely National Interest Datasets (such as those that provide registers of businesses, services or assets; or those that record activity in key areas of public expenditure), others may be less immediately obvious but will become clear candidates over time. This element of the Framework is designed to allow such evolution, with public scrutiny in each case.

Datasets with national interest characteristics may be identified: in prior research or program evaluation within the relevant sector; through use of datasets with comparable features or circumstances in other sectors or overseas; or may be inferred from the interest in or demand for access to particular datasets over an extended period.

To be a valuable resource, the suite of National Interest Datasets must extend beyond the low hanging fruit of spatial data and aggregated activity data to include access to de-identified datasets that are integral to service delivery and decision making, as well as key privately held datasets. The Framework established is intended to promote the inclusion of such data to enable its broader use, while not dis-incentivising data collection and value adding activities.

Extensive community and stakeholder consultation is expected to be an important aspect of the dataset designation process. To enhance community consultation in the process and ensure ongoing input, a deliberative forum — a parliamentary committee, in our current thinking — could be established to review nominations made and the level of access granted, and make proposals for future designations. A mechanism of this kind would ensure that there is detailed consideration of the existing pool of datasets from which nominated sets can be drawn. It would also open to public scrutiny arguments against designation.

Why designate a dataset?

The Commission's recommendations provide a Framework for public and private datasets to be nominated and designated as National Interest Datasets (NIDs). For those datasets that are so designated, all restrictions to access and use that are contained in a variety of national and state/territory legislation, and in other program-specific policies, would be replaced by new access and use arrangements under the proposed Data Sharing and Release Act (enabled in the states and territories by the Australian Government's powers under section 51(v) of the Constitution). This would ensure ongoing dataset curation as a national asset, substantially streamline access to the dataset and, where relevant, enable linkage to other datasets.

The process to designate datasets as NIDs should be open to the states and territories and to private sector entities, to allow them to similarly benefit from having their data curated (to the extent that is in the public interest) and accessible under the new Act, and more readily allow linkage of their data with other datasets. Where states and territories opt in to have datasets designated as NIDs, separate state/territory legislation may be required to enable release of data held by state government bodies and some unincorporated entities.

Having a system for identifying and funding the ongoing maintenance of national interest data assets would help create consensus and cooperation between sectors and between the Commonwealth and the states and territories. This would build on existing work at COAG to identify a spine of essential public sector assets.

A listing of all NIDs that have been publicly released or are potentially available to share, the relevant dataset custodian and ARA for that dataset, and a contact point, would be included on a central website, such as on data.gov.au. This would enable potential users of these to know of the dataset's existence and how to gain access to it.

What would access look like?

Under the Commission’s Framework, a valuable suite of datasets would be developed. At the discretion of the ARA, these datasets would either be released publicly or, alternatively, shared with all Australian and state and territory government agencies and other trusted users, under rules to be developed by the NDC.

The approach represents a marked expansion in data access in Australia that would provide significant opportunities for research and innovative market development and improve delivery of public services.

In contrast to existing arrangements for access to significant datasets, the approach recommended aims to expand:

- the availability and quality of that group of high value datasets — in the private and public sector, across all levels of government — that are of national interest
- the range of data users that would be considered ‘trusted’ to access de-identified data
- the types of uses to which the data can be put — by allowing unlimited use of data that is not about individual persons or businesses, and approved access to de-identified data to be ongoing (not project specific) and limiting use only where the risks of re-identification cannot be effectively managed.

The special case of higher risk data shared with trusted users

Some publicly funded data that identifies individual persons or businesses is already shared in a very limited way with trusted users within government and/or the research community. This data is typically used for targeted program and product/service delivery, regulatory compliance, and for research (such as rare medical conditions) where there are very small populations involved.

The current process is, however, costly to data custodians, those who endeavour to gain access to the data, and also for the public, who ultimately fund the activities for which the data is used. Depending on the particular dataset, access requests (even from within the same government) can require separate and duplicative agreement of multiple dataset owners, custodians and stewards, integration units, ethics committees, other advisory bodies, and the individuals about whom the information was collected. Each policy and approval step is intended to ensure privacy and confidentiality are maintained, but in combination they create major obstacles to data access.

The Commission recommends streamlining access to identifiable data within and between Australian governments, and for the limited range of other trusted users with which such data is shared. It is intended that identifiable data that could be shared would include both that collected by, or on behalf of, government agencies and that collected through publicly funded research bodies and projects.

In addition, there is a need for the research community to put its house in order when it comes to data sharing. Just as government data custodians should consider that they hold data not solely for their own purposes but in the public interest on behalf of citizens, so too should the data of publicly funded research be available beyond the initial researchers. And where it is not, much better justification and record-keeping is needed, to at least enable other researchers to learn what data has already been collected.

Key features of the recommended approach include that:

- access be granted on a *project-specific* basis to approved personnel in either Australian or state/territory government agencies and to approved researchers
- projects for which the data could be used would be subject to a pre-approved list of public interest purposes and require approval of the ARA and, if relevant, an ethics committee (but not the data custodian)
 - the NDC, in consultation with dataset custodians, would develop and provide the list of approved purposes for the dataset to an ARA
 - if a project does not satisfy the list of approved purposes, the applicant would be able to apply to the ARA for special access
- existing exceptions on the need to obtain consent of individuals for use of personal information for health and medical research purposes, would be extended to cover public interest research more generally
- access would occur in a specified secure computing environment with output from the dataset reviewed by an automated process prior to project completion to ensure confidentiality requirements have been satisfied
- responsibility for appropriate use of datasets would rest with trusted users, with clear and significant consequences for any breach of this trust, to provide additional incentives for maintaining the security of data and appropriate data use.

To the extent that pre-approved data purposes cover the range of uses to which data will be reasonably applied, transferring final approval from initial data custodians over to the ARA and ethics committees will substantially streamline access to data.

Making other data readily available to all

Governments have proven to be poorly equipped in understanding consumer and business demand for data and in making non-personal datasets, and those that are not genuinely commercial-in-confidence, widely available.

While the reasons for governments' inability to derive value from their data holdings may at times be understandable — governments are not entrepreneurial nor would we necessarily want them to be — they are at other times disappointing. Risk aversion is not desirable where it results in the public interest being poorly served.

There needs to be a shift in emphasis from only releasing data on request for particular projects, toward actively pushing data out in a coordinated way. In principle, all datasets in fields where there are burgeoning opportunities and capability would be opened up and released, as resources and sectoral demand allow.

This would mean that all data that is non-confidential and not related to individual people or businesses would be routinely available for use by governments, consumers, businesses and the research community. This includes information that, while it may identify individuals, is already in the public domain in some form (property ownership, for example). A realistic assessment of the risks associated with public release of identifiable information that is already public in a less accessible form, should be undertaken.

Such an approach has the potential to make a marked difference to the range and volume of data available for decision making, innovative activity and improved service delivery in the community.

The challenges in achieving this should not be underestimated. There is a very real culture of risk aversion and risk avoidance in the public sector when it comes to data release. Changing this will require strong and consistent leadership, backed up by policies that clearly spell out objectives and expectations. In addition, releasing data (and the costs associated with that) presents questions regarding whether and how access should be charged and the extent to which government agencies should or should not seek to add value to datasets.

Simply put, the Commission recommends that governments should adopt a zero or low (genuinely marginal) cost approach for data release, consistent with increasing access and the achievement of public interest benefits related to that. The exception is when a clear and compelling commercial demand and proposition for a particular dataset exists, in which case a more commercial or market-driven approach to pricing may be desirable and consistent with public interest objectives. Similarly, any value adding should only be undertaken where there is a demonstrable willingness and capacity to pay on the part of the user, and where a number of criteria related to agency capability are able to be met.

In other words, it is expected that most released datasets would be curated with metadata and necessary provenance, but otherwise be free, plain vanilla, fit for release and timely.

While release of public sector data would be the focus of governments, it is anticipated that once governments start to more actively push data out, this will encourage private entities to do likewise and to profit from doing so. That is, across the economy the value will shift from being embodied in the data itself, to being derived from the clever analysis and use of data.

Implementation

The staging of reforms will be important — with major reforms such as these, establishment of the Framework in full should be viewed from the outset as a project in need of a firm implementation plan.

Negotiation and consultation will be required with state and territory governments (as significant data holders); some parts of the private sector (for similar reasons); and with sectoral groups where NIDs are sectoral in nature (health or education).

Reforming public sector data access is a strong first step, and Australia's governments need to make significant changes if open government agendas are to catch up with those in competing economies.

The Comprehensive Right for consumers proposed will need discussion and information campaigns to help people use it to the fullest extent. Firms affected will need to be involved. Although some may have doubts, even in the course of this Inquiry's first six months some key firms (a number of banks) have recognised that allowing consumers to discover and realise benefits from their data is a key driver in building community trust.

Implementation of the proposed arrangements for providing access to NIDs and to identifiable data may take slightly longer to progress, will require consideration of transition arrangements for existing data users, and require the finer details with regard to key roles, mechanisms and technical approaches to be bedded down and based on a process of extensive consultation.

Technological developments and solutions may ensure that, over time, some of the changes required to facilitate improved and secure sharing and release of data will become easier to attain.

Yet this is not advice that implies de-coupling of parts of the Framework. In a project that aims to create new opportunity for both public and private benefit, each element supports the others. Removal of one (or more) will imbalance the opportunities and reduce the prospect for broad community acceptance.

Findings and recommendations

Addressing specific impediments to *public* sector data access

DRAFT FINDING 3.1

Australia's provision of open access to data is below comparable countries with similar governance structures — including the United States and the United Kingdom. There remains considerable scope to improve the range of datasets published (and, correspondingly, the diversity of agencies and research bodies publicly releasing data) and the usability of open data portals.

DRAFT FINDING 3.2

Data integration in some jurisdictions (particularly Western Australia and New South Wales) has made good progress in some fields, but highlights a lack of action in equivalent fields at both national and state/territory level, and reveals the large unmet potential in data integration.

DRAFT FINDING 3.3

Despite recent statements in favour of greater openness, many areas of Australia's public sector continue to exhibit a reluctance to share or release data. The entrenched culture of risk aversion, reinforced by a range of policy requirements and approval processes, greatly inhibits data discovery, analysis and use.

DRAFT RECOMMENDATION 3.1

All Australian Government agencies should create comprehensive, easy to access data registers (listing both data that is available and that which is not) by 1 October 2017 and publish these registers on data.gov.au.

States and territories should create an equivalent model where one does not exist and in all cases should make registers comprehensive. These should in turn be linked to data.gov.au.

The central agencies responsible for data should:

- set measurable objectives, consistent with best practice, for ensuring that available data and metadata are catalogued and searchable, in a machine-readable format
- improve accessibility of data for potential data users.

Limited exceptions for high sensitivity datasets should apply. Where they do, a notice indicating certain unspecified datasets that have been assessed as Not Available should be published by the responsible department of state, on the relevant registry.

DRAFT FINDING 3.4

There is a clear public interest in having research-oriented data widely available to trusted researchers in a timely manner. A corresponding presumption that it be released needs to be balanced against a number of potentially competing interests, including:

- the need for the researcher to benefit from their own research
- interests in commercialisation of research — for example, if the research was partly privately funded
- specific legislative or ethics approval restrictions
- privacy or confidentiality considerations
- capacity to provide access through secure sharing environments, where privacy or confidentiality considerations cannot be managed to enable the release of data.

DRAFT RECOMMENDATION 3.2

Publicly funded entities, including the Australian Research Council, should publish up-to-date registers of data holdings, including metadata, that they fund or hold.

Publication of summary descriptions of datasets held by funded researchers but not released, and an explanation of why these datasets are not available, are also essential and would provide far greater transparency about what is being funded by taxpayers but withheld.

Addressing specific impediments to *private* sector data access

DRAFT RECOMMENDATION 4.1

The Australian Government should adopt a minimum target for voluntary participation in Comprehensive Credit Reporting of 40% of accounts. If this target is not achieved by 30 June 2017, the Government should circulate draft legislation to impose mandatory reporting by 31 December 2017.

DRAFT RECOMMENDATION 4.2

All Australian governments entering into contracts with the private sector, which involve the creation of datasets in the course of delivering public services, should assess the strategic significance and public interest value of the data prior to contracting. Where data is assessed to be valuable, governments should retain the right to access or purchase that data in machine readable form and apply any analysis that is within the public interest.

The conundrum of personal data

DRAFT FINDING 5.1

The boundaries of personal information are constantly shifting, in response to technological advances and community expectations. The legal definition of personal information, contained in the *Privacy Act 1988* (Cth), gives rise to uncertainty. This uncertainty will only increase in future, as new technology continues to emerge.

DRAFT RECOMMENDATION 5.1

In conjunction with the Australian Bureau of Statistics and other agencies with data de-identification expertise, the Office of the Australian Information Commissioner should develop and publish practical guidance on best practice de-identification processes.

To increase confidence in data de-identification, the Office of the Australian Information Commissioner should be afforded the power to certify, at its discretion, when entities are using best practice de-identification processes.

DRAFT RECOMMENDATION 5.2

The *Privacy Act 1988* (Cth) exceptions that allow access to identifiable information for the purposes of health and medical research without seeking individuals' agreement, should be expanded to apply to all research that is determined to be in the public interest.

The Office of the Australian Information Commissioner should develop and publish guidance on the inputs required to establish a public interest case.

DRAFT FINDING 5.2

A wide range of more than 500 secrecy and privacy provisions in Commonwealth legislation plus other policies and guidelines impose considerable limitations on the availability and use of identifiable data. While some may remain valid, they are rarely reviewed or modified. Many will no longer be fit for purpose.

Incremental change to data management frameworks is unlikely to be either effective or timely, given the proliferation of these restrictions.

DRAFT FINDING 5.3

Although parts of the government view community expectations as a factor that limits the use of data, reliable surveys have shown that most individuals believe sharing personal information between government departments can be beneficial, and indeed is occurring without damage.

However, individuals expect to remain in control of who data on them is shared with.

DRAFT FINDING 5.4

Large volumes of identifiable information are already published online by individuals or collected by various organisations, with or without explicit consent.

In this context, the incremental risk of allowing increased access to formerly identifiable data by public and private sector organisations, using security protocols and trusted user models, is likely very small.

Breaches of personal data, often enabled by individuals' unwary approach to offering data, are largely dominated by malicious or criminal activity. By comparison, breaches due to sharing or release are far fewer in number and reach.

DRAFT RECOMMENDATION 5.3

The Australian Government should abolish its requirement to destroy linked datasets and statistical linkage keys at the completion of researchers' data integration projects.

Data custodians should use a risk-based approach to determine how to enable ongoing use of linked datasets. The value added to original datasets by researchers should be retained and available to other dataset users.

INFORMATION REQUEST

The Commission seeks further views on the most practical ways to ensure improvements to linked datasets are available for subsequent dataset uses.

DRAFT RECOMMENDATION 5.4

To streamline approval processes for data access, the Australian Government should:

- issue clear guidance to data custodians on their rights and responsibilities, ensuring that requests for data access are dealt with in a timely and efficient manner;
- require that data custodians report annually on their handling of requests for data access;
- prioritise funding to academic institutions that implement mutual recognition of approvals issued by accredited human research ethics committees.

State and territory governments should mirror these approaches to enable use of data for jurisdictional comparisons and cross-jurisdiction research.

DRAFT RECOMMENDATION 5.5

In light of the Australian Government's commitment to open data, additional qualified entities should be accredited to undertake data linkage.

State-based data linkage units should be able to apply for accreditation by the National Data Custodian (Draft Recommendation 9.5) to allow them to link Australian Government data, and the intention of 'open by default' should apply to these exchanges.

Making data more useful

DRAFT FINDING 6.1

The lack of public release and data sharing between government entities has contributed to fragmentation and duplication of data collection activities. This not only wastes public and private sector resources but also places a larger than necessary reporting burden on individuals and organisations.

DRAFT RECOMMENDATION 6.1

Government agencies should adopt and implement data management standards to support increased data availability and use as part of their implementation of the Australian Government's Public Data Policy Statement.

These standards should:

- be published on agency websites
- be adopted in consultation with data users and draw on existing standards where feasible
- recognise sector-specific differences in data collection and use
- support the sharing of data across Australian governments and agencies
- enable all digitally collected data and metadata to be available in commonly used machine readable formats (that are relevant to the function or field in which the data was collected or will likely be most commonly used), including where relevant and authorised, for machine to machine interaction.

Policy documents outlining the standards and how they will be implemented should be available in draft form for consultation by the end of 2017, with standards implemented by the end of 2020.

Agencies that do not adopt agreed sector-specific standards would be noted as not fully implementing the Australian Government's Public Data Policy and would be required to work under a nominated Accredited Release Authority (Draft Recommendation 9.6) to improve the quality of their data holdings.

DRAFT RECOMMENDATION 6.2

The private sector is likely to be best placed to determine sector-specific standards for its data sharing between firms, where required by reforms proposed under the new data Framework.

In the event that voluntary approaches to determining standards and data quality do not emerge or adequately enable data access and transfer (including where sought by consumers), governments should facilitate this, when deemed to be in the public interest to do so.

INFORMATION REQUEST

The Commission seeks more information on the benefits and costs of a legislative presumption in favour of providing data in an application programming interface (API) format, specifically:

- In which sectors would consumers benefit from being able to access data in an API format?*
 - What are the main costs and barriers to implementing APIs?*
-

DRAFT FINDING 6.2

Data standards should aim to ensure that the content produced is usable by those who seek access to their own data. To achieve this, available data needs to be published in machine readable and commonly used formats that are relevant to the function or field in which the data was originally collected or will likely be most commonly used.

Valuing and pricing data

DRAFT RECOMMENDATION 2.1

In determining datasets for public release, a central government agency with policy responsibility for data should maintain a system whereby all Australian governments' agencies, researchers and the private sector can, on an ongoing basis, nominate datasets or combinations of datasets for public release, with the initial priority being the release of high value, in-demand datasets.

A list of requested datasets should be published. Decisions regarding dataset release or otherwise, and access arrangements, should be transparent. Agencies should provide explanations where priority datasets are not subsequently released on legitimate grounds. Where there are not legitimate reasons for withholding requested data, remedial action should be undertaken by the Australian Government's central data agency to assist agencies to satisfy data requests.

Existing government data initiatives, such as data.gov.au, should be leveraged as part of this system.

DRAFT RECOMMENDATION 7.1

Beyond achieving a 'fit for release' standard (Draft Recommendation 6.1), government agencies should only value add to data if there is an identified public interest purpose for the agency to undertake additional value adding, or:

- the agency can perform the value adding more efficiently than either any private sector entities or end users of the data; and
- users have a demonstrable willingness to pay for the value added product; and
- the agency has the capability and capacity in-house or under existing contract; and
- the information technology upgrade risk is assessed and found to be small.

DRAFT FINDING 7.1

There is no single pricing approach that could act as a model for guiding public sector data release decisions. The identification by agencies of the grounds for undertaking each release will have a direct bearing on the choice of price approach. Cost recovery, long considered to be the default option in the public sector, is only one of a range of approaches and not necessarily to be preferred.

DRAFT RECOMMENDATION 7.2

The pricing of public sector datasets to the research community for public interest purposes should be the subject of an independent review.

DRAFT RECOMMENDATION 7.3

Minimally processed public sector datasets should be made freely available or priced at marginal cost of release.

Where there is a demand and public interest rationale for value-added datasets, agencies should adopt a cost recovery pricing approach. Further, they should experiment with lower prices to gauge the price sensitivity of demand, with a view to sustaining lower prices if demand proves to be reasonably price sensitive.

DRAFT RECOMMENDATION 7.4

For datasets determined through the central data agency's public request process (Draft Recommendation 2.1) to be of high value and have a strong public interest case for their release, agencies should be funded for this purpose. Funding should be limited and supplemental in nature, payable only in the event that agencies make the datasets available through release or sharing.

Aside from this additional funding, normal budgetary processes should apply for all agencies' activities related to their data holdings.

Fundamental reform is needed

DRAFT FINDING 8.1

It is important governments and businesses maintain a social licence for their collection and use of data. This can be built through enhancement of consumer rights, genuine safeguards, transparency, and effective management of risk. Community trust and acceptance will be vital for the implementation of any reforms to Australia's data infrastructure.

DRAFT FINDING 8.2

There is no shared vision amongst public sector data holders in Australia on how to consistently deliver widespread data sharing and release. The community — current and future — is entitled to expect such a vision. Comprehensive reform of Australia's data infrastructure is needed to signal that permission is granted for active data sharing and release and that data infrastructure and assets are a priority. Reforms should be underpinned by:

- clear and consistent leadership
- transparency and accountability for release and risk management
- reformed policies and legislation
- institutional change.

DRAFT FINDING 8.3

By applying a risk-based approach to data access, government agencies can establish a sound basis for where further risk mitigation effort is necessary and for moving early to the sharing or release of low risk data, while building and retaining the trust and confidence of users and the wider community.

DRAFT RECOMMENDATION 9.1

The Australian Government should introduce a definition of consumer data that includes:

- personal information, as defined in the *Privacy Act 1988* (Cth)
- all files posted online by the consumer
- all data derived from consumers' online transactions or Internet-connected activity
- other data associated with transactions or activity that is relevant to the transfer of data to a nominated third party.

Data that is transformed to a significant extent, such that it is demonstrably not able to be re-identified as being related to an individual, should not, for the purposes of defining and implementing any Comprehensive Right, be defined as consumer data.

The definition of 'consumer data' should be provided as part of a new Act regarding data sharing and release (Draft Recommendation 9.11). Given the need for this definition to have broad applicability, it should also be included within the *Acts Interpretation Act 1901* (Cth). Consequential amendments to other Commonwealth legislation would ensure harmonisation across federal laws.

INFORMATION REQUEST

Further views are sought on the effects of providing access to consumer data, as defined. In particular, views are sought on the potential creation of incentives for deliberate de-identification of data holdings to avoid providing access, and whether effective and low cost remedies to such behaviour could be introduced.

DRAFT RECOMMENDATION 9.2

Individuals should have a Comprehensive Right to access digitally held data about themselves. This access right would give the individual a right to:

- continuing shared access with the data holder
- access the data provided directly by the individual, collected in the course of other actions (and including administrative datasets), or created by others, for example through re-identification
- request edits or corrections for reasons of accuracy
- be informed about the intention to disclose or sell data about them to third parties
- appeal automated decisions
- direct data holders to copy data in machine-readable form, either to the individual or to a nominated third party.

Individuals should also have the right, at any time, to opt out of a data collection process, subject to a number of exceptions. Exceptions would include data collected or used as:

- a condition of continued delivery of a product or service to the individual
- necessary to satisfy legal obligations or legal claims
- necessary for a specific public interest purpose (including archival)
- part of a National Interest Dataset (as defined in Draft Recommendation 9.4).

The right to cease collection would not give individuals the capacity to prevent use of data collected on the individual up to the point of such cessation.

INFORMATION REQUEST

The Commission seeks views on what methods of disclosure would be most likely to result in consumers making a meaningful choice about how their personal information is being used, and how these disclosure requirements might best be implemented.

DRAFT RECOMMENDATION 9.3

The Australian Government should provide for broad oversight and complaints handling functions within a reformed framework for individual data access. Key roles should be accorded to the Australian Competition and Consumer Commission (ACCC) the Office of the Australian Information Commissioner (OAIC), and to existing industry ombudsmen.

Any charging regimes, policies or practices introduced to address costs associated with data access, editing or transferability should be transparent and reasonable. The ACCC should be responsible for monitoring and assessing the reasonableness of charges applied. The ACCC, supported by state and territory Fair Trading Offices, should also educate and advise consumers on their new rights in regard to data access and collection.

For specified datasets (such as in banking) the relevant ombudsman scheme would need to be expanded to deal with disputes.

INFORMATION REQUEST

The Commission seeks further views on datasets that are of national interest and that could feasibly be designated as such under the process proposed.

DRAFT RECOMMENDATION 9.4

The Australian Government, in consultation with state and territory governments, should establish a process whereby public and private datasets are able to be nominated and designated as National Interest Datasets (NIDs).

Datasets (across the public and private sector) designated as NIDs would satisfy an underlying public interest test and their release would be likely to generate significant community-wide net benefits. Designation would occur via a disallowable instrument on the recommendation of the National Data Custodian.

NIDs that contain non-sensitive data should be immediately released. Those NIDs that include data on individuals would be available initially only to trusted users and in a manner that retains the privacy of individuals and/or the confidentiality of individual businesses. The in-principle aim should be for these de-identified datasets to be publicly released in time.

The process to designate datasets as being of national interest should be open to the states and territories in order to cover linked datasets, with negotiations undertaken to achieve this.

For community confidence, consideration should be given to use of a deliberative forum, such as a parliamentary committee, to take community input on and review nominations made, and to make proposals for future designations.

DRAFT RECOMMENDATION 9.5

The Australian Government should establish an Office of the National Data Custodian, as a new function within the Government to have overall responsibility for the implementation of data management policy.

Specifically, the National Data Custodian (NDC) would have responsibility for broad oversight and monitoring of Australia's data system, recommending the designation of National Interest Datasets, and accrediting Release Authorities and trusted users within the reformed data system.

DRAFT RECOMMENDATION 9.6

Selected Australian and state/territory government agencies should be accredited as Release Authorities by the National Data Custodian. In considering applications for accreditation, the National Data Custodian should consult a wide range of parties and ensure Accredited Release Authorities (ARAs) have sectoral expertise. The current model used by the National Statistical Service for appointing data linkage authorities should be considered in developing a model upon which to base this process.

ARAs will be responsible for:

- deciding (in consultation with initial data custodians) whether a dataset is available for public release or limited sharing with trusted users
- collating, curating and ensuring the timely updating of National Interest Datasets.

ARAs will also perform an important advisory role in regard to technical matters, both to government, and to the broader community of data custodians and data users.

DRAFT RECOMMENDATION 9.7

Trusted users should be accredited by the National Data Custodian for access to those National Interest Datasets (NIDs) that are not publicly released. Trusted users should be drawn from a wide range of potential entities, including: all Australian Government and state and territory government agencies; all Australian universities; and other entities (be they corporations, not-for-profit organisations or research bodies) that are covered by privacy legislation.

The default position should be that someone from one of these organisations would be approved for access unless the National Data Custodian transparently specifies a reason, on consideration, of why this should not occur.

For trusted users of NIDs, trusted user status should provide an ongoing access arrangement, with few restrictions on what could be done with the data. Trusted user status for NIDs should cease when the user leaves the approved organisation or be suspended if a breach occurs by any other trusted user in that same organisation and/or working on the same project.

INFORMATION REQUEST

The Commission seeks further views on the establishment of a Parliamentary Committee to take community input on possible National Interest Datasets, to review nominations made, and make proposals for future designations. Views are also sought on practical alternatives.

DRAFT RECOMMENDATION 9.8

Arrangements for access by trusted users to identifiable data held in the public sector and by publicly funded research bodies should be streamlined and expanded by the Australian Government. The National Data Custodian should be given responsibility to:

- develop, in consultation with data custodians, a list of pre-approved uses for a dataset, and make decisions on access to data for projects not consistent with the pre-approved uses list
- grant, on an approved project-specific basis, trusted user access to personnel from a range of potential entities, including: all Australian Government and state and territory government agencies; all Australian universities; and other entities (be they corporations, not-for-profit organisations or research bodies) that:
 - are covered by privacy legislation
 - have the necessary governance structures and processes in place to address the risks of inappropriate data use associated with particular datasets, including access to secure computing infrastructure.

Access would be granted for the life of the specific approved project. Trusted user status for use of identifiable data would cease when the user leaves the approved organisation; a project is completed; or if a breach occurs in that same organisation and/or project.

DRAFT RECOMMENDATION 9.9

Public research funding should be prioritised on the basis of progress made by research institutions in making their researchers' data widely available to other trusted researchers on conclusion of research projects.

DRAFT RECOMMENDATION 9.10

All non-sensitive public sector data should be released, consistent with release priorities and as resources allow, with curation, provision of metadata and adherence to agreed standards resourced as specified in Draft Recommendation 7.4. A realistic assessment of the risks associated with public release of identifiable information that is already public in a less accessible form, should be undertaken by all governments.

Data that could be used for program or agency performance management purposes should not be withheld from release.

DRAFT RECOMMENDATION 9.11

The Australian Government should introduce a *Data Sharing and Release Act* which includes the following:

- Provisions requiring government agencies to share and release data with other government agencies and requiring sharing between government agencies and other sectors.
 - These provisions would operate regardless of all restrictions on data sharing or release contained in other legislation, policies or guidelines.
 - The provisions may be waived in limited exceptional circumstances, and the Act should specify what these circumstances are.
- Strengthened provisions on access to data by individuals, including rights to access and edit data about them, a right to have data copied and transferred, and a right to request that collection cease.
- Provisions establishing the Framework for the governance of Comprehensive Rights of consumers, access to National Interest Datasets, approval of trusted users, and accreditation processes for Release Authorities.