



**Australian Government**  
**Department of Industry,  
Innovation and Science**

Secretary

Mr Peter Harris  
Chairman  
Productivity Commission  
GPO Box 1428  
Canberra City ACT 2601

Dear Mr Harris

Thank you for the invitation to provide a submission to the Productivity Commission inquiry into Data Availability and Use.

The Department of Industry, Innovation and Science is committed to improving the availability and use of data in accordance with the Government's Public Data Policy. To this end, the department and IP Australia welcome the opportunity to make a submission to the Productivity Commission on this important issue.

Please find attached our submission for the Commission's consideration.

Yours sincerely

Glenys Beauchamp

August 2016

# Department of Industry, Innovation and Science submission to the Productivity Commission inquiry into Data Availability and Use

## 1. The broad settings for data availability and use

The Department of Industry, Innovation and Science is embracing the opportunities presented by the attention being given to increasing the availability and improving the use of data.

The Public Sector Data Management Project and the Public Data Policy Statement provided a good opportunity to start a new conversation about the value of public data and to establish some positive short term actions, but more is required.

The department considers that a broader strategy will be required to access the full benefits available. This strategy would establish enforceable standards for data use in the public sector, preferably across all levels of government. It would also clearly articulate expectations relating to private sector data and build the social licence required to use data well. Achieving these objectives will require champions in the public and private sectors and possibly also require broad-ranging legislative action.

Rather than seeing itself simply as a 'creator' or 'user' of data, the department sees itself as an enabler of more effective policies, programmes and new knowledge through enhanced data availability and use, whether that value happens to be created by the department, the broader public sector, businesses or the not for profit sector.

## 2. The benefits of increasing the availability and use of public sector data

### 2.1. Potential benefits

Data is one of the few economic resources that is abundant rather than scarce, and can be used and re-used by multiple parties. This means that there is an economic rationale for increasing access to, and use of, this increasing resource to benefit the community.<sup>1</sup> It follows that public data offers a host of benefits by virtue of being made available. These potential benefits include but are not limited to:

- **Generating new knowledge** about how our economy and its various components work.
- **More effective policy development and programme delivery**, by leveraging data assets across government to identify the policies and programmes that have the most impact and achieving improved services to and interactions with users of various public services.
- **Enhancing both public sector and private sector innovation**, as access to and use of open data will enable the creation of new services (both in the public and private sectors) and the growth of industries that specialise in adding value to previously unavailable data.
- **Enhancing public sector efficiency**, as access to and use of open data will allow the public sector to find ways to do more with less.
- **Increasing the transparency and accountability of government** which will increase public trust in government.
- **Improving public sector collaboration** by decreasing informational barriers between organisations both public and private.

---

<sup>1</sup> Gruen N, Houghton J, and Tooth R, (2014), 'Open data for business: How open data can help achieve the g20 growth target', Lateral Economics, [https://www.omidyar.com/sites/default/files/file\\_archive/insights/ON%20Report\\_061114\\_FNL.pdf](https://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf)

The department holds rich data assets such as data on interventions to improve the competitiveness, innovation and entrepreneurship of business and industries; R&D and research commercialisation; energy efficiency; energy and resource production; macro-level industry statistics and long-term industry trends. When combined with private data holdings, increased data sharing also has the potential to boost collaboration between industry, government and the research community, a noted area of weakness in the Australian innovation system.

Through better access to data and linked data, researchers will be able to generate new knowledge. This knowledge will contribute to a better understanding of the Australian economy.

From a policy perspective, data-enabled research can create and inform the evidence base of what works and what does not. The community as a whole is expected to further benefit from openly available data through an increase in evidence-based policy and policy transparency. The department currently uses data to improve services and information to key stakeholders and business, advise ministers and support decision making within the department. Liberating data further will improve the relationship between the government and the community.

## **2.2. Linking public sector datasets**

Further liberation of public sector data from restrictions on sharing and use creates opportunities for more datasets to be linked. Many public sector datasets are of minimal use in isolation, and their value increases exponentially when they are linked to other datasets. These linked datasets generate new opportunities for public research and government policy development, as linked public sector data can provide a picture of the economy and society that is richer, more accurate, and more detailed than any individual dataset could in isolation.

These benefits can be magnified when linking complementary datasets from different agencies, enabling innovative “wraparound” programmes and services, where assistance from multiple agencies is provided in a coherent, coordinated fashion to more effectively address an issue. Information drawn from these linked datasets has the potential to significantly enhance the efficiency and quality of government services, enhance accountability and transparency, and drive innovation. Effective information sharing across agencies can deliver better services to individuals and businesses (for example to achieve ‘tell us once’ services, personalisation and pre-filling of forms). However, there are a range of factors that work against such outcomes, as discussed in section 3 below.

In addition to these benefits, there are many public sector datasets that cannot be reproduced easily, if at all, in the private domain and would be of potential use to researchers, academia, and commercial and entrepreneurial entities. In particular, there is great potential for the linking of public sector data with privately held data. Such datasets could be used by the private sector to find and develop new markets, effectively guide investment and mitigate risk, and thus contribute to economic growth.

## **2.3. Examples of linking in public sector data**

### Expanded Analytical Business Longitudinal Database (EABLD)

The Expanded Analytical Business Longitudinal Database (EABLD) provides an excellent example of the benefits that result from linking datasets and increasing the availability of public sector data. The EABLD was developed by the Australian Bureau of Statistics (ABS) in collaboration with the Department of Industry, Innovation and Science. The initial version integrated administrative data from the Australian

Taxation Office (ATO) with collected survey data from the Business Characteristics Survey (BCS), Economic Activity Survey, Business Research and Development Survey and intellectual property data.

A study was recently carried out using the EABLD to explore employment growth which concluded that start-ups drive employment growth in the economy, and yet start-up activity is in long term decline. The findings from research such as this, drawn from large, linked datasets, influenced the development of the National Innovation and Science Agenda and will inform future policy and programme development, leading to community benefits that are the direct result of the open sharing of data held in the public sector with the rest of the community.

The purpose of the EABLD is to develop an enduring firm-level statistical asset that will increase the capacity of the public sector and the research community to undertake firm-level analysis of micro-economic drivers of performance, competitiveness, and productivity, and thus provide a solid evidence base for future policy development and programme evaluation. As the first comprehensive dataset of all Australian businesses that includes detailed financial data, it will enable the comparison of business performance between businesses that participate in government programmes with those that do not. Programme evaluations conducted using the EABLD will allow better identification of programme features and characteristics that have the most impact, leading to a more targeted expenditure of government funds.

Access to the EABLD is significantly constrained by the legislation governing ABS and ATO data. Currently, mechanisms have been devised to facilitate access by Australian Public Service staff, although these are less than satisfactory. Trials are in progress to enable academic researchers to access the EABLD to conduct research projects such as on the relationships between business characteristics, innovation, productivity, job creation, entrepreneurship and performance. The department is working with the ABS and Data61 on a project to scope future enhancements to the EABLD, including identifying new data for inclusion and mechanisms for improving access for researchers.

#### Australia National Data Service

The Australian National Data Service (ANDS), a partnership between Monash University, the Australian National University and the CSIRO and funded by the Australian Government, aims to build the value of Australia's research data assets. This initiative has improved access to and use of government and research organisation datasets, facilitating further research and innovation. Many of Australia's key science partners and competitors – including the US, UK, EU and China, have adopted, or are in the process of implementing, policies to make their science and research systems more open. Australia has much to benefit from building on the work of existing initiatives such as ANDS by further expanding access to data and research findings.

#### IP Australia

The administrative data contained in the Intellectual Property Government Open Data (IPGOD) dataset operated by IP Australia provides another example of the benefits of linking data. IPGOD provides valuable insights into who uses the Intellectual Property system, and where and how it is used. By linking this data with Australian Business Numbers and Goods and Services Tax registration data, IP Australia was able to carry out firm-level analysis on who uses IP rights in Australia. Other examples include IP Australia's work using:

- National Survey of Research Commercialisation data, which links patent data with research undertaken by Australian publicly funded research organisations;
- Pharmaceutical Benefits Scheme data, linking data on pharmaceuticals that are listed on the PBS with patent data.

These examples highlight innovative uses of open data which can be used to inform policy debate, as well as reduce administrative burden by having the data collated and indexed.

#### VANguard and the Unique Student Identifier

VANguard is a whole of government program that provides authentication services to more than 42 Australian and state government agencies. Built on highly-available and accredited infrastructure, VANguard allows agencies to focus on delivering digital services without having to deal with the cost and complexity of building in-house authentication services. VANguard utilises business data held by the Australian Business Register (within the Australian Taxation Office) to deliver 105 million AUSkey authentications each year.

An example of VANguard in use is the Unique Student Identifier (USI) initiative. The Department of Education utilises VANguard to authenticate Registered Training Organisations (RTOs) accessing student information via the USI registry system. The USI system allows RTOs to login - using AUSkey and authenticating with VANguard - to register, report against and manage over 4 million registered vocational education students in Australia.

### **3. Factors preventing government agencies from making their data available**

When the availability of public sector data is discussed, it is worth considering that there are different audiences with whom it is possible to share data. These users can be described as:

- internal (within a department or agency);
- external but within the Commonwealth (i.e. to other Commonwealth departments and agencies);
- trusted external (i.e. academics, consultants, other Australian governments, international intergovernmental organisations); and
- unrestricted.

There are a variety of impediments to sharing to each of these audiences. These include cultural barriers, obligations of confidentiality in programme guidelines and legislative barriers.

#### **3.1. Cultural barriers**

Over decades the department (and its predecessors) has, as with other departments, developed data handling practices designed to minimise the risk posed by inappropriate disclosure and use of data. There are no legislative restrictions on the sharing of the identifiable unit record level data for the vast majority of the department's programme data holdings. However, in some cases, restrictive data handling practices has become integrated into day-to-day activities, becoming indistinguishable from those that are legislated. This process, combined with a risk-averse culture, has resulted in non-legislated practices that appeared as valid as any legislative requirement.

When working with administrative data, it is important to understand how the data was collected and how the nature of that source shapes the information and conclusions that can be drawn from it. For

example, much of the administrative data available from business assistance programmes derives from businesses that were eligible to participate in the programme; thus there are a number of characteristics that these businesses can be expected to share that would be different to the overall business population. It is common for this knowledge to be tacit rather than documented, meaning that often only people who routinely work with the datasets are familiar with the range of limitations that apply to this data. When the keepers of such data are approached by others wishing to use it, awareness of the limitations of the data coupled with the magnitude of work required to adequately document these creates a bias against release.

Furthermore, even when documentation is available, some keepers of data do not trust others to understand the limitations of data, and fear any potential misinterpretation. The risk of potential misinterpretation is high for some of the department's programme data holdings, as stakeholders may interpret results from analytical work in ways that the data stewards may not. Mitigating this issue requires strategic oversight, coordination and good data governance to support analytical work to be undertaken using departmental data holdings.

There is a mirrored risk of misinterpretation of data by data stewards and analysts, due to oversight or an unduly narrow focus. Examples may include the evaluation of a policy's impact without reference to an equivalent untreated population, or evaluation of customer preferences using an unrepresentative sample. Improved public access to linked datasets informing these decisions mitigates the risk of misinterpretation, whilst improving public confidence in government and evidence-based policy.

### **3.2. Obligations of Confidentiality and Legislative Barriers**

In addition to the cultural barriers to sharing data, in some cases obligations of confidentiality created by programme guidelines (including Ministerial guidelines) and other documents, and genuine legislative restrictions operate as barriers to sharing.

#### Confidentiality

Under the common law duty of confidence, the department may only use or disclose information obtained from programme participants within the boundaries of authorised use. Authorised use will include the primary purpose for which the information was provided, purposes which would be reasonably expected by the provider of the information, and any other use or disclosure to which the provider consents to either at the time of providing the information or subsequently. The existence and nature of this consent is commonly determined from notice of how the department will use or further disclose the information in programme guidelines (and other documents such as application forms, information guides and other disclaimers) at the time the information was given, as well as the circumstances in which the information was provided. The use or disclosure of confidential information beyond authorised use may constitute a breach of confidence, which may provide grounds for legal action. As such, programme guidelines should not create obligations of confidentiality, and unnecessarily restrict the use of information obtained from programme participants, where it is not required. In addition, internal data sharing rules should not be more restrictive than the department's obligations in the programme guidelines.

#### Legislative Barriers

The majority of the department's programmes have no legislative restriction on sharing of the identifiable unit record level data, however there are a small number of restricted programmes.

The following are examples of legislation that restrict the sharing of data for all or key parts of a number of programmes in the department, as well as the Australian Business Register (ABR). The secrecy provision of the *Pooled Development Funds Act 1992 (s.71.)* (PDF Act) restricts sharing of protected information within the department unless the information is to be disclosed for the purposes of the PDF Act or the *Venture Capital Act 2002*, (VC Act) or where the disclosure is made in the performance of that person's functions under the PDF Act. The confidentiality provisions in Divisions 355B and 355C of Schedule 1 of the *Taxation Administration Act 1953* (TA Act) make it an offence for staff to further disclose confidential information provided to them by the ATO except for the purpose for which that information was given – and that this remains the case even if the entity involved consents to the disclosure. A consequence of these provisions is that it is not possible to provide a complete picture of departmental assistance to a particular firm if they have accessed the R&D Tax Incentive, since the quantum of revenue forgone (and thus the amount of benefit obtained by the firm) is protected by the TA Act and cannot be disclosed.

A further example of the issues that arise from the current wording of the TA Act is that it prevents data obtained by the ATO from being incorporated into the ABR, which is administered by the ATO. Additionally, non-sensitive information (such as turnover and employee numbers expressed as ranges) in the ABR from ATO records would be a valuable resource for research and policy development, and would avoid the need to repeatedly ask businesses for this information. However, the TA Act as well as the others mentioned above prevent this data from being shared, and are particularly problematic as there is no process described in the legislation that would allow authorisation of data sharing. Where these are specified, it can be possible to address some restrictions without resorting to legislative amendments. For example, the *Industry Research and Development Act 1986* restricts the use of protected information, however it also allows the Secretary to designate people to whom information to be disclosed. In order to enable information from R&D tax measures to be included in the departmental customer relationship management (CRM) system, the Secretary authorised the disclosure of this information to departmental officers who used this system. However, if restrictions on sharing programme data governed by acts such as the PDF Act, VC Act and TA Act are to be lessened, legislative amendment is required.

The ABR, being a complete list of all Australian businesses, both active and former, is a very valuable public data asset. Data in the ABR is divided into public and private data, with the public data accessible by the public through the ABN Lookup service offered by the Department of Industry, Innovation and Science. However, there are valuable data items (most notably, industry codes) which are designated as private data, whose use is restricted by the *A New Tax System (Australian Business Number) Act 1999* (ABN Act). Given that most of the private data in the ABR for a business can be uncovered through a web search, it is recommended that the operation of the privacy and confidentiality provisions of the ABR be reviewed to ensure that they are still fit for purpose. The unlocking of the information contained in the ABR for public use has the potential to be of enormous benefit for both the public and private sectors.

In order to share and publish data for public benefit, it is necessary to know what data can be legally shared or published. It is also necessary to know whether publishing and sharing is allowed at the identifiable unit record level or only in confidentialised or aggregated forms. The sharing rules for the department's programme data are complex. Approximately two-fifths of the department's programme datasets operate under 20 separate pieces of legislation. Additionally, each separate legislative instrument differs in describing the treatment of data, and capacity for interpretation exists. Given this complexity it is not surprising that interpretations of the legislation can vary, with a strong preference to

minimise the risk of breaching legislation through data disclosure, and thus supporting the risk averse culture towards data sharing noted earlier.

Until recently, the most common mechanism for sharing data between government entities was a Memorandum of Understanding. These were often unnecessarily complicated and time consuming – it could take several years and multiple memoranda to establish data sharing arrangements between government entities. Further, these agreements were not legally binding and could incur significant cost in policy, project and legal officer time. This process has been superseded by the Guidance on Data Sharing for Australian Government Entities. This is a good example of an administrative change that, once fully implemented, will facilitate the availability and use of data and data sharing.

### **3.3. Data Confidentialisation**

Interwoven amongst the previously discussed barriers to data sharing is the issue of data confidentialisation - this is a further practical and legal consideration in making public sector data open. Staff regularly publish aggregated figures derived from confidential datasets, which is permitted provided the data has been adequately confidentialised. The process of making data confidential involves removing data items that would allow for the identification of individuals or organisations, while simultaneously maintaining the structure and integrity of the dataset. Currently, there is no standard description of how this process should be done across the entire public service.

Any useful standard would specify how data should be prepared for publication, and would provide a checklist for staff releasing data to third parties or publishing in the public domain. Such a standard would not only mitigate the risk of unauthorised confidential data disclosure, and thereby allay fears of a community backlash to wider use or release of personal information, but would improve the quality of published data, give confidence to those undertaking the work that they are doing it correctly, and reduce the amount of time spent making the data confidential by streamlining the process. All of these outcomes are necessary if public sector data is to be routinely shared.

As datasets are integrated there are increased risks of 'spontaneous recognition', whereby confidentialised datasets may still yield sufficient details for a researcher to identify the entity about whom data relates to. This could be an accidental discovery by an astute individual, or a sought-after discovery by a determined individual. The important point is not to try to prevent this risk at all, but to find the right balance between managing the risk of spontaneous recognition whilst preserving the value of the data. The five safes model, promoted by organisations such as the ABS and the Australian Data Archive, is part of the answer, recognising that trusted researchers will have little interest in acting upon accidental recognition and that trusted output arrangements allow for the checking of products derived from access to data.<sup>2</sup> These arrangements can be bolstered by having appropriate procedures put in place quickly in the case of an event of spontaneous recognition.

### **3.4. Sharing Commercial-in-Confidence data**

Currently, there is no general legislation governing the sharing of commercial-in-confidence material. With the exception of legislative secrecy provisions applying to particular datasets, the concept of commercial-in-confidence is derived from common law. While the case law regarding when a duty of confidence will arise may be sufficiently clear to identify whether information collected in a particular

---

<sup>2</sup> Australian Bureau of Statistics, [Transforming Statistics for the Future, Feb 2016](#), ABS Cat. No. 1015.0, 2 February 2016. Accessed 13 July 2016.



instance is confidential, there are a variety of approaches taken across different programmes and different government agencies. Accordingly, there is some degree of confusion around how commercial-in-confidence data should be handled, which in turn leads to a risk minimising and restrictive approach to sharing.

The lack of clarity around what commercial-in-confidence protections entail has impacts across government, and indeed, in the private sector as well. If the sharing and publishing of data subject to these protections is to be facilitated, then some sort of more uniform treatment of commercial-in-confidence data would be beneficial.

One option to achieve this could be an act governing commercial-in-confidence data that would clearly define what information is protected and what can be disclosed within and between government agencies without breaching the duties of confidence attached to that information, in a similar manner to how the *Privacy Act 1988* does for personal information. Such an act could take into consideration that, typically, commercial-in-confidence data has the potential to be damaging for only a short time, and that restricting the sharing of such data in the long term may be a barrier with no purpose.

An alternative approach could be to have a central Commonwealth policy regarding the collection and sharing of programme data which can be applied across the Commonwealth. This policy could outline the extent to which programme data is to be collected on a confidential basis and expectations regarding data sharing between Commonwealth agencies. This central approach could then be adopted as standard across programme guidelines, which would result in the same type of confidentiality obligations arising across data collected in relation to different programmes and different circumstances, and a shared understanding both from programme participants and across Commonwealth agencies regarding the extent to which this data can be shared, at least amongst government agencies.

### **3.5. Security**

Public sector data that is made openly available is either de-identified or aggregated to maintain the privacy of the owners of the data. De-identified or highly aggregated data is not very useful for in-depth analysis that can be carried out by internal research partners (within a department or agency and within the Commonwealth) or trusted external research partners (i.e. academics, consultants, other Australian governments, international intergovernmental organisations). These partners usually require access to the data in its raw form meaning that some of this data will be disaggregated, sensitive and/or confidential. There are security, time and access barriers to sharing such data.

While de-identified or aggregated data can easily be shared with partners and the public on data sharing platforms such as [data.gov.au](http://data.gov.au), the cloud environment is increasingly becoming identified as a secure platform to share disaggregated, sensitive public sector data with internal and external research partners. In these cloud environments data can be encrypted and user access limited to trusted researchers. There will need to be increased support for, and advocacy of, standard permission schemes, such as the Australian Signals Directorate (ASD) Certification which awards certification to a range of cloud service providers for specified cloud services. This could help departments understand and mitigate security risks, and create a more flexible and technology adaptive data sharing environment for sensitive data.

### **3.6. Data quality**

Other than the previously discussed cultural and legislative barriers to sharing, the quality of the data also has an impact. Many public sector datasets require a significant amount of ‘cleaning’ work to be undertaken to correct errors and duplication before they can be analysed, shared with a wider audience or linked with other datasets. The process of making a dataset ready for analytics and linking demands a large investment of resources – the data must be cleaned, tidied, standardised, confidentialised, etc. Since the value of data increases rapidly as connections between separate datasets are made, and an important goal of open data is to realise the potential of data through linking, it is important that the appropriate resources are invested in improving data quality.

### **3.7. Technological Barriers**

Legislation is not always able to keep up with technology and new ways of working with data. This can cause inefficiencies and inconsistencies in how agencies release and deal with their data products. For instance, legislation might specify the mandatory release of data through one medium whilst restricting it through another. For example, IP Australia is mandated to publish intellectual property right application data through its register, but is then restricted in providing that same data in other formats. A possible solution to this issue might be a legislative interpretation that considers data that is released by government agencies as falling under a creative commons license by default.

There is also the challenge of dealing with data that is contained in legacy systems. Even though these systems can be modernised, there are a number of technical challenges associated with the effective sharing of such data. These challenges include: an inability to link data tables; standardisation issues between data systems; gaps in metadata availability; and inconsistent storage formats. All of these can lead to data quality issues and difficulties in automating data provision.

In addition to this, some degree of data error is inevitable when dealing with legacy systems. In order to deal with this, IP Australia, an organisation that holds data on legacy systems, has adopted an approach of releasing data along with a programme of constant improvement to the data. This ensures that at any time the data is the ‘best available’. This approach has allowed IP Australia to work collaboratively with its data users to uncover and resolve data issues as part of the data release process, and to take an iterative approach of including more data offerings as underlying issues are addressed. This constant release governance approach has also motivated a programme of data validation and integration within IP Australia, and is illustrative of the unanticipated systemic benefits of flexible open data governance and policies.

### **3.8. Motivation and incentives**

There are currently fewer reward mechanisms in place to encourage data sharing than there are disincentives. A notable disincentive is that making data openly available involves additional work related to preparing and documenting data for release and publication and this requires additional resources that can be hard to find in a resource constrained environment. There is also a lack of experience regarding how to make data available and maintaining it using relevant data infrastructures (e.g. data.gov.au). A further disincentive to data sharing is the penalty regime that exists for data breaches. By taking into account unintentional breaches and reducing penalties for this, a culture that more readily engages with data sharing may result.

A regime that provides more corporate recognition and rewards for the sharing of data would help to mitigate the previous concerns and disincentives. Towards this end, the role of ‘data custodian’ should be a recognised one within government agencies. Such a role would enable collaboration between Chief

Information Officers and operational level custodians from across government, and would allow for a common training platform for custodians across government.

GovHack is an example of an initiative that promotes the benefits of public data sharing and encourages individuals to become involved. It is a high profile volunteer initiative with active government participation that works in conjunction with data.gov.au to promote awareness of data in the broader research and public communities. The department and IP Australia have offered multiple prizes at the GovHack competitions and found that it generated innovative and fresh perspectives on what can be achieved with openly available government data. For example, IP Australia adapted and implemented on a larger scale a multi-prize winning hack which was created using IP data from IPGOD, to create Intellectual Property Neural Open-Data Visualisation and Analytics (IPNOVA). By bringing the real life benefits of open data to the foreground, GovHack increases the motivation for agencies and the Australian data community to do more with public data assets.

#### **4. Department of Industry, Innovation and Science perspectives on data sharing**

The department has mandated internal data sharing where there are no legislative restrictions, and is taking action to improve its data management and is making its data more accessible. As discussed above, the department recognises that making our data holdings available for research and collaboration will assist it to meet the Government and the public's expectations of effective, continually improving service delivery, and policies informed by the best available data.

Consequently, the department created the Data Management and Analytics Taskforce to build its data capability and to advance the mandate to share its data holdings internally, and externally subject to appropriate controls to ensure any public release is in compliance with legislative obligations and with due consideration of commercial-in-confidence protections. A principle objective of the Taskforce is to address the cultural barriers and clarify legislative barriers. The Taskforce will work towards removing these cultural barriers by concentrating multi-disciplinary resources on these data sharing challenges, and by providing sufficient and recurrent funding for staff to do this work effectively. The Taskforce is also responsible for actioning the Australian Government's Public Data Policy Statement within the department.

It is important to note that the department does not believe it should release every dataset in its possession for which it is technically possible to do so. Many datasets would be of little use if released into the public domain as they contain too few businesses for statistical analysis, and their release would contribute little to analysis, research or government transparency. As discussed above, a dataset often requires a large amount of cleaning and standardising work to be done on it before it can be released to the public. An element of cost/benefit practicality must therefore be considered when making public sector data open. In a resource-constrained environment, data publishing efforts must be focused in the first instance on datasets that will deliver returns to the public and community commensurate with the amount of resources required to prepare them for release.

Notwithstanding the department's determination to make its data holdings more accessible, it remains committed to upholding the privacy and confidentiality undertakings that it has made to its clients, as well as those arising from legislation. The department has systems in place to ensure that information is protected, so that businesses are not detrimentally affected as a consequence of providing information.

#### **5. Summary**

In summary, there are currently a number of barriers to the effective sharing of public data internally. In the department's experience, the main barriers are cultural practices based on risk aversion, and uncertainty regarding obligations of confidentiality, which together contribute to a bias against data sharing. These barriers are being addressed internally within the Department of Industry, Innovation and Science. On a broader level, there are genuine legislative barriers to sharing data, both internally and externally, as well as barriers such as uncertainty around how commercial-in-confidence data should be handled and consistency in data collection approaches, and these will require a whole of government response. Overall, the department is very supportive of increasing the availability of public sector data, and believes that the community stands to greatly benefit from improved service and policy delivery. For this reason it is important for barriers to be reduced, minimised or even eliminated if possible.