



12 August 2016

Data Availability and Use
Productivity Commission
GPO Box 1428
CANBERRA CITY ACT 2601

By online lodgement only: <http://www.pc.gov.au/inquiries/current/data-access/>

Dear Commissioners,

Re: Inquiry into Data Availability and Use

The Quantium Group Pty Ltd ('Quantium') welcomes the opportunity to provide comments on the Productivity Commission's *Data Availability and Use* Issues Paper (the "Inquiry").

Background to Quantium

Quantium is one of Australia's leading data analytics firms. We have one of the most powerful big data clusters in Australia, a unique privacy management system and a privacy compliant data ecosystem consisting of large de-identified consumer data sets (including data from Woolworths, National Australia Bank ("NAB"), Foxtel and CoreLogic).

Through a combination of rich data assets, verifiably effective privacy management, cutting edge technology and advanced analytics, we provide our clients with valuable inferences about interests and preferences of segments of the Australian population. Our business provides a combination of consulting services, customer profiling and predictive modelling to assist our clients to make key business decisions, including more granular marketing decisions, while respecting individuals' right to privacy.

Commencing operations in October 2002, Quantium now has offices in Sydney, Brisbane and Melbourne, overseas in India, New Zealand and South Africa, and presently employs over 480 people worldwide.

Our core proposition is a unique set of data partnerships that interwork with our privacy management. Complemented by our media activation partners Facebook, News Corp and oOh! Media, we can seamlessly activate digital media campaigns and measure the effectiveness of the message.



Drawing on our experience in commercialising and combining public and private sector data, we comment below on some issues raised in the Inquiry: specifically, the attributes of a valuable dataset, the challenges that parties face when working with data, and valuable uses of public sector data in a data-enabled economy.

Data as a resource

Data is a key economic resource for Australia, now and into the future. Production and collection of data is ubiquitous in modern society. The daily choices we make – from the food we eat, where and how we travel, to the way we consume and disseminate information – are captured at various points and by numerous parties. Individuals produce and voluntarily share data in nearly every aspect of their lives. The recent explosion of new technologies such as smartphones and wearable devices, combined with new analytical know-how presents the opportunity to understand our lives in greater detail than ever before. Key to Australia capitalising on this opportunity is continuing to educate consumers about the public benefits that flow from the collection, use and value of data, within a regulatory framework that is easy to navigate and which strikes the right balance of societal benefit and each individual's right to privacy, including a right to know how, when and where personal information about an individual is being collected, used and disclosed.

Q: What characteristics define high value data sets?

High value data sets (both public and private) are:

- **Timely**, and collected as close as possible to real-time and on a regular basis;
- **Granular**, with as much detail as possible;
- **Comprehensive**, representing a cross section of the population;
- contain only **Accurate and Real** information (as opposed to self-reported); and
- **Linkable** to other data sets through the presence of identifiers.

Challenges

Q: What are the main factors currently stopping government agencies from making their public data available?

Any release of personal information requires a measured and considered balance between protecting the privacy of individuals and deriving utility from the data. Government agencies are rightly conscious that once data sets are released into the public domain, they can no longer be controlled. As a consequence, many agencies do not wish to put any data into the public domain for fear of making the wrong decision. In a world that is fast moving and with technology ever-changing, there are a number of factors that impede agencies from making their data available. These include:

- the need to manage privacy, including compliance with privacy protocols, conduct of privacy impact assessments and safeguarding de-identified data sets from unacceptable risk of re-identification of individuals;



- keeping abreast of technology changes;
- managing opt-out requests and individual exceptions;
- storage of large data sets and the security risks attached to that storage;
- the risk of unintended or malicious use of data;
- information security and managing risks of misuse of data;
- attracting people to the agency who are equipped to manage data issues; and
- understanding the benefits that will flow to the agency.

The **comprehensive** and **granular** nature of government data sets makes their release difficult. At an individual level, individuals may be prone to re-identification, despite best efforts to de-identify them. Conversely, efforts to mitigate these risks through aggregation and smoothing may reduce many of the benefits of these data. Data sets ultimately are of most utility when used at the most granular level, however public data releases often err on the side of caution thus limiting its potential use.

We acknowledge that public sector agencies at federal, state and local government levels already generate, curate and publish a large amount of data relating to their activities. These data sets exist as a collection of related but disjointed data that reflect the multitude of source government agencies. Initiatives such as www.data.gov.au, the Population Health Research Network (PHRN) and state based Open Data policies have begun opening and joining data for use by the general public.

These initiatives have been successful in allowing academic researchers access to public data. In isolation and at lower levels of granularity, the process of accessing public data is relatively straightforward. As the use cases become more sophisticated hence requiring more data sets, the process of locating the relevant data custodians and gaining approval for use and linkage of data becomes increasingly difficult and slow. This is especially the case where the data sets cross multiple jurisdictions. The length of time (and associated cost) to obtain data may prove to be too great a hurdle for commercial users.

The approach the private sector has taken is to divide data releases and exchanges into public and trusted zones. Within the public zone, data is made safe through high levels of de-identification. For example, data points may be aggregated into cells of certain sizes or outliers removed. Often these releases fall under Open Data Commons licences. More granular data is made available to trusted parties who have implemented and adhere to the necessary security, privacy and commercial controls to ensure privacy compliant handling of de-identified information without any real risk of re-identification of any individual to whom de-identified data relates. Within the trusted de-identification zone of Quantum's data ecosystem, data matched by us includes data sets from the Woolworths, NAB, Foxtel and CoreLogic.

We suggest that a two tier system might also be applied to the controlled release of public sector data – namely a public tier and a trusted tier. The trusted tier would only be available for use by trusted parties who have implemented and verifiably and reliably adhere to the necessary security, privacy and commercial controls.



Challenges in distribution

How can the fragmented public data sets be brought together in a streamlined and accessible manner?

It is our recommendation that:

1. Existing indexes such as www.data.gov.au be expanded to include data sets that do not currently fall under the Open Data licences, so that data custodians can be readily found and approvals sought.
2. Enduring linkages (ready-made individual level linkages between multiple data sets) within the PHRN and other linkage authorities be expanded within public sector and incorporate key private sector data sets. This will enable:
 - a. efficiencies through the reuse of common linkages;
 - b. single approval processes for groups of related data sets, for example, all related data within a linked set could have pre-approved use cases; and
 - c. development of on-going applications that utilise linked data.
3. A framework and standard be established for the safe exchange and storage of public sector data sets outside of the existing data laboratories allowing the increasing demand for data sets to be distributed across the market and hence enable further uses of public sector data without the burden of additional high availability infrastructure to the public sector.

Q: How can the sharing and linking of private sector data be improved in Australia?

The business sector should remain able to commercialise data relating to customers and internet users on terms determined by businesses, within the significant constraints already imposed through various laws including laws relating to privacy of individuals and data protection, surveillance and tracking technologies, and competition and consumer laws. Data is a valuable asset, in which businesses now invest heavily. Outside the current legislative reporting requirements and operation of competition laws, we would not support any mandatory regime for private sector data to be shared or otherwise made available to other private sector entities or generally.

For those businesses in the private sector that are willing to share and link their data, we believe there are ways in which this can be achieved. For example, as part of our engagements, we facilitate linkages between our partnership data sets and other private sector data sets. The ease of establishing this linkage varies from party to party, usually due to differences in:

1. the encryption protocol to be used, which ranges from raw data fields to salt encryption; and
2. the data security standards at the destination party, which varies depending on each organisation but is largely based on ISO27000 standards.



Quantum has developed a standard protocol that streamlines this process to facilitate these exchanges, which sets out:

- the roles and responsibilities of all parties involved;
- the security requirements for the handling and storage of the data; and
- steps for the exchange.

We believe wider adoption of such protocols, in conjunction with other industry platforms will allow more use cases to develop both in the public and private sector.

Quantum is one of the founding members of Data Governance Australia (DGA), an industry body which will formally launch in September 2016. Its founding members are highly credentialed and motivated in the space of data governance, use and access. Chaired by former ACCC chair Graeme Samuel AC, DGA's founding membership base includes companies such as Woolworths, NAB, Qantas, Westpac and Coles. Its charter is to develop and communicate a self-regulated governance framework for the data industry, and to provide an industry platform for liaising with the Privacy Commissioner and Government. DGA is ideally placed to play a significant role with Government and relevant stakeholders in developing any framework to improve the sharing and linking of private sector data. We believe that an industry self-regulation model is more responsive to evolving consumer and business concerns than a legislative framework which, by its nature, must endeavour to anticipate uncertain and unpredictable disruptive developments in business and technology.

Q: What are the reasonable concerns that businesses have about increasing the availability of their data?

Most businesses addressing the increased availability of their data (either in a trusted exchange or in the public domain) are concerned as to the possibility of misuse, the consequential loss of trust of customers, users, staff and business partners and loss of competitive advantage. Maintaining consumer trust is critical to garnering support for the release of data from the public and private sector. There have been a number of examples in the US where 'de-identified' data sets have been re-identified, leading to consumer scepticism regarding the security of private sector data sets.¹ The recent movement encouraging Australians to misreport information in the 2016 census, in response to privacy and security concerns, is an example of how such scepticism can erode both the quality and integrity of the data sets.

¹ See, eg, Latanya Sweeney, 'Matching Known Patients to Health Records in Washington State Data' (2013) Data Privacy Lab <<http://dataprivacylab.org/projects/wa/1089-1.pdf>>

See also Anthony Tockar, 'Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset' on *Neustar Research* (15 September 2014) <<https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/>>

See also Arvind Narayanan and Vitaly Shmatikov, 'Robust De-anonymization of Large Sparse Datasets' (Paper presented at IEEE Symposium on Security and Privacy, The Claremont Resort, Oakland, USA, 19 May 2008) 111-125 <http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf?utm_source=dataflog&utm_medium=ref&utm_campaign=dataflog>



Legislative controls will be unable to keep pace given the rapid growth in rich, multi-dimensional data sets (such as Twitter) and advances in computing power. The combination of data and technology will bring with it new and innovative ways to derive value but it will challenge any rigid legislative framework for the collection, use and dissemination of data. Any frameworks devised must be adaptable to ensure high levels of public confidence. We acknowledge the role that a clear framework will play in assisting business and the community in understanding rules around collection, sharing and manipulation of data. In view of the nature of the assets and the inevitable technological developments that will continue to drive its value, this is best managed in a self-regulated environment.

A data-enabled economy

Q: What benefits would the community derive from increasing the availability and use of public sector data?

Realising the value of public sector data sets

Data availability is only the first part of the journey to realising community benefits. All data have to be treated, enhanced and transformed to allow information and insights to be presented in a meaningful and timely manner. This process requires both specialist knowledge as well as infrastructure that may not be available to the wider community.

The examples where public sector data have been utilised for community benefit both domestic and overseas have been the fruits of organised entities funded by government or private organisations. Notable examples using public sector data sets include:

- **The National Map** (www.nationalmap.gov.au) is a publically funded project where public sector data sets were centralised and transformed by National Information Communications Technology Australia (NICTA) in order to allow the community to easily access and visualise information from public sector agencies.
- **MySchool** (<https://www.myschool.edu.au/>) provides information about all schools in Australia, including the number of students, revenue, and academic performance. Releasing this data enables:
 - parents to make an informed choice about their child's education;
 - educators to understand and benchmark their school's performance; and
 - not-for-profit organisations to understand where to focus their limited resources to maximise the return back to the wider community.

Privately funded collaborations using both public and private sector data sets have led to some of the most innovative data-driven products in Australia. Examples of these initiatives from Quantum include:

- **PeopleLikeU** (www.peoplelikeu.com.au) is a publicly available tool enabling consumers to, for the first time, compare their spending habits with other Australians. The community is able to understand their



spending habits relative to the population and explore new recommendations. It utilises the ABS census data as well as de-identified transactional customer records from UBank and NAB.



- The **NAB Online Retail Sales Index** (<http://business.nab.com.au/tag/online-retail-sales-index/>) allows consumers and businesses to track the Australian online economy on a quarterly basis. The index is based on a combination of data from the Australian Bureau of Statistics (including the census and retail sales data) and NAB transactional data.
- The **NAB Charitable Giving Index** (<http://business.nab.com.au/tag/charitable-giving-index/>) provides insights relating to charitable donations made in Australian economy, which has been especially valuable to the not-for-profit sector.

The road ahead

Australia's data journey is well underway with the necessary components of source data, skilled labour and the technology converging. There are still many important factors to address, especially the balance between maintaining privacy and maximising utility, improving and increasing channels for data distribution and encouraging investment in developing infrastructure and skills to realise the value of data sets. We believe that private-public sector partnerships can expedite this journey.



We appreciate this opportunity to present this submission to the Inquiry and would be pleased to provide whatever further assistance the Commission considers helpful. Please do not hesitate to contact Ben Ashton, Quantum's General Counsel, _____, should you wish to discuss this further.

Yours sincerely,

.....
ADAM DRIUSSI
Chief Executive Officer
The Quantum Group Pty Ltd