

15 December 2016

Mr Peter Harris AO
Chairman, Productivity Commission
GPO Box 1428
Canberra City ACT 2601
By email: data.access@pc.gov.au

Dear Mr Harris

**Productivity Commission – Public Inquiry into Data Availability and Use –
UWA response to draft report**

The University of Western Australia's overarching strategic research agenda is encapsulated under the premise of "Data Intensive Discovery". This is based on the belief that future knowledge will increasingly be driven by our ability to produce, navigate and interrogate data in increasingly sophisticated ways.

The University is generally supportive of the Draft Report and of the introduction of new legislation that will facilitate access to, and use of, both public and private datasets in the research environment. However, we would like to take the opportunity to make the following comments:

1. Past the tipping point of curatable data

A general point about the Draft Report is that although it presents a substantial move forward in our appreciation of the value of data and what analysis of that data could do for the economy, for society and the health and wellbeing of the population, it still focusses on processes that are administration-heavy and largely manual. The reason that technologies like IBM Watson are ushering in an age of cognitive computing is that there is a fundamental recognition that we have gone beyond the point at which we can reasonably handle data being produced manually. The idea of data "curation" belongs to a time when normal people could be expected to understand and know how every bit of data an organisation handled is structured and related to every other bit of data. This point has passed. Organisations, Governments, and Health Services have already gone past the point of being able to even document a fraction of the data they produce.

A new hospital being built today (such as the Fiona Stanley Hospital or Perth Children's Hospital) will have more than 30,000 sensors producing terabytes of data whose interconnections can't be fathomed by anything other than an AI. Curation is not possible and the term suggests that the curators have an *a priori* knowledge of how the data is going to be interpreted and analysed, which of course they do not.

The consequence of this is that manual processes will simply fail because they will involve too much cost and effort and the benefits will not necessarily be seen by the organisation bearing the cost of producing the data.

2. Data and the cognitive age

The University believes that the only way in which data sharing and re-use can be promoted is to reduce the administrative overheads involved in collecting, preparing, documenting and publishing data. Technology can help with this in that tools can be created to automate this process as much as possible - especially in the de-identification/anonymisation part of the process.

Technologies like IBM's Watson will be critical because they can operate on unstructured data. These technologies will be required to allow for the interrogation or preparation of the data in order to be able to make any use of it.

A consequence of not keeping the costs of production of data low is that the cost of the data to use in research will be prohibitive. Already, the analysis and re-uses of historic data from a social media site like Twitter is of the order of \$20,000 for fewer than 5 queries. This puts it out of the reach of PhD students and most other researchers. This fate will face most if not all data if cost-recovery is encouraged, especially when the costs to be recovered are excessive because of the processes involved.

3. Providing secure environments to produce more freedom of data access

The Centre D'accès Sécurisé aux Données (CASD) in France looked at the issue of secure environments from a different perspective by convincing the French Government to entrust a large array of datasets to the researchers it supports through the use of an ultra-secure access mechanism. Similar in concept to the Sax Institute's SURE, the CASD takes the security up a level with biometric identification and hardware-based access. Their philosophy of data access has changed as a result of this environment. Their view is to provide researchers with whatever data they need for however long they need it for. This lifts a restriction of researchers needing to know exactly what they are looking for before they have had an opportunity to search the data. Essentially, if you have a secure environment that you are certain controls what researchers can do with data, you can give them more freedom with what data they access and for how long.

4. Consumers and personal big data

The University believes that we have passed the point at which the amount and range of data being collected by devices, phones and software is meaningful to individual consumers. Consumer level genetic testing will now provide 1GB of data relating to 50 million points of genetic information for an individual. A wearable tracker produces potentially 15,000 data points a day that will include heart rate, steps, and potentially location and other metrics. An environment sensor in a house will again provide similar amounts of information.

As far as consumer data goes, it is doubtful that a consumer would be able to enumerate all of the different systems and devices collecting information. Nor would they be in a position to assess the potential privacy impact of that data even if they had expressed interest in the details in the first place. There is also the question of whether they would have any understanding of the possible privacy leakage of data through secondary interpretation of the data. For example, a temperature/humidity sensor in a room measuring the room's temperature could quite easily be used to work out what a person was doing in that room and at what time from the potentially minute changes in environment that they themselves could bring about. The introduction of the Internet of Things into a home will radically change what software and systems know about the people inhabiting it. This is a good thing from the perspective of the opportunities of assisting ageing in place and assisting those with disabilities, but also has a flip side with the degree to which it records private and even intimate information.

Yours sincerely



Professor Robyn Owens
Deputy Vice-Chancellor (Research)