



**Australian
Privacy
Foundation**

<http://www.privacy.org.au>

Secretary@privacy.org.au

<http://www.privacy.org.au/About/Contacts.html>

26 July 2016

Mr T. Pilgrim
Privacy Commissioner
Office of the Australian Information Commissioner

Dear Timothy

Re: Submission: 'Big Data and the APPs'

APF welcomes the drafting of guidance for government agencies and corporations in relation to big data matters.

APF notes, however, that it would have been helpful to have received notification of the consultation process through the Consumer Privacy Network (CPN), rather than having to depend on its own environmental scanning processes to detect the opportunity.

There are several aspects of the draft of May 2016 that we believe need to be addressed prior to finalisation.

1. The document should avoid buying into the hype

In several passages, the document accepts the assurances of big data spruikers:

p.2 "big data analytics has changed the way ..."
rather than, for example,
"it is claimed that big data analytics is changing the way ..."

p.3 "big data has come about from a fundamental shift ..."
rather than
"proponents argue that big data involves a fundamental shift ..."

The techniques are not revolutionary, but merely an ongoing evolution, following the long-running thread of management information systems, decision support systems and data mining.

p.3 "big data has changed the ways entities use data to ..."
rather than
"big data analytics offers some additional means of using data to"

p.3 "This is different to how data was analysed in the past ..."
rather than
"This augments existing analytical techniques"

The claims made by proponents that 'applying algorithms to find correlations' is somehow new are at least misleading, and arguably false.

2. Much greater emphasis is needed on the importance of quality PIAs

The very substantial public concern arising in relation to the Census indicates the vital importance of rigorous and open PIAs for big data projects, including early engagement with civil society.

3. The document should draw attention to its narrow focus on, and only on, the APPs

APF understands that the OAIC prefers to frame its guidance within the strict confines of the Privacy Act. However, we submit that the implications of this need to be communicated to the reader.

The document should make clear that a full risk assessment of big data activities needs to extend beyond these data protection aspects to encompass such concerns as procedural fairness in relation to administrative decisions, avoidance of unfair discrimination, behavioural privacy and privacy of the person.

4. The document should refer to re-identification at an earlier stage

On p.4, the text about 'De-identified personal information' could be mis-read, and lead organisations to think that de-identification is a simple matter. This would be overcome by a short 3rd para. that looks ahead to the valuable section on re-identification on pp.5-6.

e.g. 'However, the challenges that need to be overcome when undertaking de-identification are considerable, particularly in the case of rich data-sets. This is discussed in a later section'.

5. The document should distinguish generic information about the APPs from content that is specific to big data activities.

A large proportion of the text on pp. 6-23 necessarily and sensibly replays material from prior OAIC documents. It would be clearer, however, if a different type-face were used for the two categories of text, e.g. italics for the generic material.

6. The segment on big data and quality needs further development

On p.21, there is a valuable section on quality factors. However:

(a) it is important to distinguish between:

- **the quality of data** that is acquired from one source;
- the quality of data that is acquired from multiple sources; and
- the quality of any data that may be generated by the analysis process.

Each matters, and all may be different

(b) a further major issue not currently addressed is **the quality of the inferencing processes, and of any decision-making processes that apply the inferences**. One problem is the use of techniques that assume that data is on a ratio scale, when the data may be on some other scale (cardinal, ordinal, nominal). Another is the approach to handling missing data-items

(c) another major problem that needs to be addressed is **the data's semantics**, including:

- the meaning of each data-item;
- the meaning of the existence of a record in a data-collection; and
- the compatibility of the various meanings across different sources.

Yours sincerely



Kat Lane, Vice-Chair
For the Board of the Australian Privacy Foundation
0447 620 694
Kat.Lane@privacy.org.au