

Submission to the
Review of FOI laws

December 2012

To: foireview@ag.gov.au

Thankyou for the opportunity to contribute to this consultation process.

I would like to see greater access for members of the public to aggregated statistics from government databases.

1. Each agency's Information Publication Scheme should include a list of all its databases and a description of each.
2. The Freedom of Information Act should allow applicants to specify any aggregate query to run across an agency's data sets.

Statistics are an important means of control and accountability, by letting us evaluate performance and identify trends.

The problem at present is that we, as members of the public, are only told what the government wants to tell us. Agencies get to select how they would like to frame the story, and they select statistics accordingly.

For example, I might like to know more about the types and numbers of welfare recipients, and what combinations of benefits they receive. I could consult the Joint Agency submission to the parliamentary inquiry into the sufficiency of Newstart, but that only contains two bar graphs, with statistics selected to support the story the agencies want to tell.

Under the current FOI regime, I would have to request a document which contains the information I want. No such document would exist. Even if the FOI officer wanted to be helpful by bringing such a document into existence for me, they would have to spend time seeking the data and seeing is recorded and what is not. It would be difficult for me to know exactly what I want until the officer had already done the work of gathering it, only to find out I need something different.

Consider the difference if each agency published a list of the database tables it holds, the columns within those tables, and how the tables join with each other. I could see what characteristics of DHS clients are recorded, and how that data relates to data about particular welfare benefits. I could then frame a coherent, sensible request for the statistics I need in order to understand the system and its performance.

Members of the opposition, independents, and candidates for elections could all access such a facility to hold the government to account and uncover the truth for voters. It would, to some extent, alleviate one of the unfair advantages of incumbency in an election, which is privileged access to government information and the ability to selectively publish it.

Employees of the public service would also probably benefit from a better understanding of the data held by their own and other agencies.

From a civil liberties perspective, with the government collecting all manner of information about us, it would be valuable to be able to consult a published catalogue of government data sets. We would be able to find out exactly what the current nature of government surveillance and recording is, and make an informed decision about whether we are comfortable with it or would like to pursue reform.

Open access to entire data sets presents a great privacy and secrecy problem. (I am referring to the suggestion alluded to in Prof John McMillan's speech in Canberra on 29 November.) Most data sets contain confidential information. Anonymised data can sometimes be de-anonymised when combined with other data. Many data sets would have to be thoroughly checked before they could be released.

However, access to aggregate functions over government databases presents a much smaller problem. It is very easy for an agency FOI officer to determine whether a query, or the resulting data set, contains sensitive information. The bulk of such resulting data sets would simply be numbers and percentages. The biggest risk will be that the row labels would contain specific names, or that the population of a cell would be narrowed down to an identifiable person (e.g. persons earning over \$18 billion per year). However, an agency FOI would be familiar with that agency's data and could detect such problems.

The basic rule should be that an agency must perform aggregate queries over any of its data sets, unless doing so would reveal information that would otherwise be exempt under the FOI Act.