Australian Government

**Department of Industry,
Innovation and Science**

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

Dear Mr Harris

Thank you for the invitation to provide a response to the draft report of the Productivity Commission inquiry into Data Availability and Use.

The Department of Industry, Innovation and Science is supportive of many of the recommendations contained in the draft report, and remains committed to improving the availability and use of data in accordance with the Government's Public Data Policy. To this end, the department welcomes the opportunity to provide its response to the Productivity Commission on this important issue.

Please find attached our response to the draft report for the Commission's consideration.

Yours sincerely

Glenys Beauchamp

December 2016

Response to the draft report of the Productivity Commission's inquiry on Data Availability and Use

Department of Industry, Innovation and Science

The Department of Industry, Innovation and Science broadly supports the recommendations made in the draft report of the Productivity Commission's inquiry on Data Availability and Use (the Report).

This response will include both the department's view on selected recommendations from the report, as well as responding to some of the areas where the Report requests additional information.

Data sharing legislation

The department supports the intent of the proposed Data Sharing and Release Act, particularly a requirement for government agencies to share and release data with other government agencies and requiring sharing between government agencies and other sectors. Agencies will require good data management, data skills and strong leadership to participate well in the environment envisioned by the Commission. Many will also need to adjust internal cultures.

The department has achieved significant benefits from such data sharing activities, but has also experienced restrictions associated with legal or cultural data sharing barriers internally and in other agencies.

National interest datasets/data integration

The department supports the designation of datasets as National Interest Datasets. This will improve the availability and accessibility of data that is in the public interest. It would also focus attention on significant datasets, encouraging their custodians to ensure that the data is well managed, of good quality, and its collection and maintenance appropriately resourced for the good of the whole community.

The designation of National Interest Datasets by a National Data Custodian is an appropriate process for their identification. The suggestion of using a deliberative forum to take community input on and review recommendations for National Interest Datasets has merit. After an initial rush it is likely that these nominations are likely to be few and infrequent. The Report recommends that a Parliamentary Committee review nominations for dataset nomination and the level of access to be granted to that dataset. This function would require transparency, subject matter expertise, and a good awareness of privacy, confidentiality and ethics concerns, from both a legal and community perspective.

Although the criteria for constituting a National Interest Dataset are yet to be fully established, it is likely that data assets such as the Australian Business Register, the Business Longitudinal Analytic Data Environment (BLADE) and the Australian Taxation Office's business tax datasets would be strong candidates for consideration as National Interest Datasets. It should be noted that BLADE is comprised of multiple subsidiary datasets from a range of sources. It is unclear how National Interest Dataset designations apply to subsidiary datasets in these circumstances.

Improving the accessibility of datasets such as the business taxation datasets holds great promise for improving the efficiency of program administration and reducing provider burden. By making these datasets more accessible, departments can obtain information needed for program administration, policy development and evaluation from these sources rather than directly from program participants.

This reuse of data simplifies program administration by eliminating the need to re-collect and quality assure data, replacing it with a trusted external source. It also reduces the compliance burden on program participants by not requiring them to provide the same information to the government multiple times. The department suggests that the Report could strengthen its arguments by giving greater prominence to the potential benefits of making data available for reuse on the efficiency of government administration, and the reduction in compliance costs for individuals and businesses.

Confidentiality and Risk

There is a lack of common language used to discuss and describe data risks and data confidentiality. This complicates the task of developing a consistent approach to these issues, as people have differing understandings about what is meant by terms such as anonymised, de-identified, confidentialised, and sensitive data. For example, in the Department of Prime Minister and Cabinet's Public Sector Data Management Report, "de-identified data" is defined as being equivalent to "anonymised data", and it makes no reference to "confidentialised data", whereas the Australian Bureau of Statistics' National Statistical Service in its confidentiality glossary states about anonymised data "This term is most commonly used to refer to data from which direct identifiers have been removed (de-identified data), but it is sometimes also used to refer to confidentialised data. To avoid confusion the more specific terms de-identified data and confidentialised data are used in the Confidentiality Information Series."

Although this may appear to be a minor distinction, the risks of re-identification are considerably higher for de-identified data than for confidentialised data. The conflation of these two types of data leads to level of risk associated with the release of a dataset being misinterpreted. De-identified data may be released in situations where confidentialisation would be required to properly protect individuals' data, by officials who believe that they are complying with approved release procedures. In order to appropriately manage the Australian Government's data holdings and avoid inappropriate data releases, it is imperative that a common terminology be agreed to across government. The risks of inappropriate use or disclosure of data are magnified by the absence of agreement on what these terms mean.

Trusted users

The Report recommends that trusted users be accredited by the National Data Custodian to enable them to access National Interest Datasets that are not publicly released. The department notes that "trusted user" as a category is problematic, as it would be more appropriate to describe trust along a spectrum. A better approach may be to develop trusted user categories, each of which is accorded a level of access commensurate with the level of trust or accountability they have. This would allow for greater flexibility in dealing with the different types of users who would have an interest in accessing National Interest Datasets. Provisions could be made to allow users to move to a more trusted category if they give enforceable undertakings regarding their use of data.

Accredited Release Authorities

The department supports the concept of Accredited Release Authorities as recommended in the Report. At this time, it is not anticipated that the department would seek to be an Accredited

Release Authority; it would be appropriate for there to only be a limited number of these, primarily agencies that routinely publish sensitive data, such as the ABS.