ODI QUEENSLAND

29 July 2016

The Open Data Institute Queensland (ODI Queensland) is pleased to provide this submission in response to some of the specific questions raised in Productivity Commission Inquiry into Data Availability and Use.

About the ODI and ODI Queensland

The ODI was co-founded in 2012 by the inventor of the web Sir Tim Berners-Lee and AI expert Sir Nigel Shadbolt to address today's global challenges using the web of data. Headquartered in the UK, the ODI has an international reach, with over 30 'nodes' including the Australian node ODI Queensland, as well as hundreds of members, thousands of people trained, and dozens of startups incubated.

We are an effective, open innovation intermediary which brings together commercial and non-commercial organisations and governments around specific sectors to address today's global challenges.

We connect, equip and inspire people around the world to innovate with data by:

- Providing leadership and helping to develop strategy
- Researching and innovating
- Developing language and shaping policy
- Giving training
- Supporting and encouraging startups
- Creating global networks
- Bringing the voice of business to government

ODI Queensland was established in December 2014, with seed funding provided by Foundation Sponsors from the private sector, research and academia. On 26 July, the Queensland Government announced a two year open data partnership with ODI Queensland to better harness the social, economic and environmental benefits of reusing and repurposing public data to solve challenges, and inform decision-making.

ODI Queensland Ltd
Queensland.theodi.org
70-72 Bowen Street, Spring Hill, Queensland 4000

Our submission in response to the certain questions raised in the issues paper:

ODI Queensland hosted a discussion forum in June specifically to review certain questions posed in the issues paper. This submission is based on feedback provided at and subsequent to the discussion forum, on our experiences in general in Australia since December 2014, and on relevant guidance and advice from the international ODI community.

ODI Queensland wishes to offer its support, experience and leadership of the global ODI community to the Australian Government.

1. **Public sector data**
   Questions on high value public sector data:

   What public sector datasets should be considered high-value data to the: business sector; research sector; academics; or the broader community?

   - Areas of higher expenditure: social services and health.
   - Core reference data, being data about the things around us that make society function. They are the things we collect information about in registers. They are the things that all the information we collect and use is really about. Some basic examples include:
     - Business data eg ASIC, ABR, ATO, (data about businesses)
     - Property data, sales, construction, planning, development.
     - Environmental data generated by public sector as part of general business operations
     - Energy and water, provisioning and usage, including alternative energy (solar, wind, etc)
     - Common boundaries (postcodes, suburbs, local government areas, state areas)
     - Property boundaries
     - Addresses
     - Transport network
     - Transport data, public transport, road and transport infrastructure usage, vehicle registration, etc
     - Place names
     - Waterways
     - Schools and academies
     - Hospitals
     - Public service facilities
   - Publicly funded data collections.
   - Data about what data the public sector has. An open catalogue around archives for example.
   - Historical data.

- Research grant data
- Our legislation - compare Australia https://www.legislation.gov.au to the UK http://www.legislation.gov.uk. By opening and making legislation machine readable you can:
    - Analyse it to determine how it may inhibit data release / sharing
    - Link the collection and release of data to a law (i.e. I can legally release this private data)
    - Optimise / simplify the law

What characteristics define high-value datasets?

- Core reference data that is
    - referenced from other information. These references are the junctions onto the byways of statistical and administrative data
    - where each item is assigned an identifier, such as a number or code, to make it easy to reference and therefore create a junction with other datasets
    - where lists of them are probably maintained through some defined processes which ensures the roads and junctions themselves get maintained
- Real time data, or near real time, or at least current data.
- Data capable to helping to understand social problems and improve social outcomes for example:
    - Health data (shared not necessarily open)
    - Social Services data (shared not necessarily open)
- Data about data collections
- Metadata about data collections
- The metrics used to describe the data itself.
- Research reports that aren't necessarily considered 'data', need to include the analysis, the underlying data. To enabling the mining of this type of data as we can mine more traditionally viewed 'data'.
- "My health record", a centralised repository of personal data. Extending such data with limited access rights to start with (for certain research purposes for example).

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

The overarching benefits case for open data is now broadly accepted internationally (in part through the activities of the ODI). The McKinsey Global Institute, in *Open Data: Unlocking Innovation and Performance with Liquid Information*, determined that open data has the potential to enable $3 trillion in additional value to the global economy. The European Union estimates the size of the open data market in the 28 European Member States and the 4 European Free Trade Association (EFTA) countries will be €55.3 billion

in 2016. The value of open data in the U.S. is accepted (for example by Vivek Kundra, former U.S. Government Chief Information Officer) as amounting to $trillions. A report by PWC predicted that the ODI's challenge series in the UK would generate social and economic impacts of up to £302 million. A 2013 review estimated the total social and economic value of open data to the UK at £6.8 billion annually.

In its report *Open for Business*, Lateral Economics estimates the potential value of open data to Australia at $64 billion, with more rigorous open data policies adding around $16 billion annually to Australia's economy. The Federal Public Data Policy Statement labels open data a 'Strategic National Resource', while the ALP proposes to publish a National Information Policy and to establish an Independent Data Council to promote an 'optimal open data culture' in Australia.

The appendices to the Australian Public Sector Data Management Project report provide a range of case studies highlighting the tangible value that can be gained through open data projects, while the ODI provides a further body of evidence and stories around the benefits being realised through open data.

An innovation intermediary like the ODI is necessary to facilitate tangible outcomes and benefits to the community.

Benefits that can be derived by the community would include:

- Providing Australian businesses the opportunity to use open data as fuel for innovation, job creation, growth and economic development, through the development of new products and services, and new business insights
- Attraction of new investment, a healthy and vibrant knowledge economy
- A tangible and visible commitment by the Australian Government to transparency, accountability and citizen engagement,
- Enabling Australians to be an informed and empowered population
- bringing research, academia and industry together with government to increase capacity and collaboration around public sector priorities and
- improving government service delivery and productivity through more effective data management practises and data sharing across agency boundaries.

Questions on collection and release of public sector data

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

- Insufficient resources
- Budget
- Lack of data literacy
- Fear of doing the wrong thing

- Fear of private / sensitive data being re-identified due to inadequate de-identification processes and lack of training. Data science and tools have developed rapidly to make this more of a concern than before.
- Lack of guidance regarding standards
- 'Knowledge is power' which is a reluctance to share, for fear of losing internal competitive edge of value proposition within the organisation. Sharing data will give others access to it and allow them to gain an advantage
- Conflicting policies and perception of insufficient leadership

How could governments use their own data collections more efficiently and effectively?

The value and potential value of data beyond its original purposes is not truly appreciated within the public sector. The public sector should encourage data collectors to think about other uses and users of their data, at the time of collection. This will mean that data is collected, stored, managed and accessed in ways to support use cases beyond its initial purpose.

It would be beneficial to consider how data may be analysed and adjust the data collection to support analysis.

Other suggested advice:

- Share data internally within other Departments / jurisdictions
- Use data themselves
- Actively remove duplicate data and discourage copying of data
- Enable other users to avoid the need for copying data by providing an API to the data with a guaranteed service level
- Improve data literacy through training and education
- Enable the data to be linked together
- Leverage open standards and open formats (balanced with practicality)
- Nominate a dedicated data person internally within each Department, someone who can help guide staff as to what data is available, how to access it, how to use it
- Agree on common standards of collection, naming, publication (preferably internationally consistent).

Should the collection, sharing and release of public sector data be standardised? What would be the benefits and costs of standardising? What would standards that are 'fit for purpose' look like?

Consistency with international standards is critical. Datasets that are common across many organisations should be standardised (for example that collected by our 700+ local councils) on collections such as:

- Budgets / expenditure

- Contracts / procurement

Datasets of common reference data should be standardised and released openly http://www.anzlic.gov.au/fsdf-themes-datasets

Standards that are fit for purpose would be:

- Developed in cooperation
- Openly published and openly licensed
- Open to contributions
  See more at https://open-stand.org/about-us/principles/

Standards would enable data to be joined up http://juds.joinedupdata.org.

The policies and regulations which cross over each other and conflict need to be reviewed to eliminate conflict.

What criteria and decision-making tools do government agencies use to decide which public sector data to make publicly available and how much processing to undertake before it is released?

- Proactively release based on the law e.g. Queensland's Right to Information legislation.
- Release based on demand - engage with potential users of data
- Improve data based on demand and requirement to comply with adopted open standard.
- Perform data quality checks in source system to enable provision of data quality statement along with published data.
- Release the data as it is and provide an opportunity for feedback
- Be able to describe the risks, fitness for purpose information in the metadata
- Policies conflict
- Review legislation that may inhibit release of data

What specific government initiatives (whether Australian Government, state, territory or local government, or overseas jurisdictions) have been particularly effective in improving data access and use?

A number of Queensland government agencies have utlised the ODI's Open Data Maturity Assessment Model. This has enabled them to discover their strengths and weaknesses, baseline their maturity as a contributor of open data, and then benchmark their score on a global comparative. They have used the Model to then identify areas for improvement to optimise progression and receive practical recommendations to help achieve their goals.

The greater the organisation's open data maturity the greater the usability of the data they publish and share. A mature open data publisher is a mature:

- Collector through good quality business process (and putting effort into how you collect data, has a corresponding positive impact on the business process behind the collection)
- Custodian of data - they see data as a valuable, reusable resource to themselves but also to others; they expend appropriate effort in managing and storing it, that befits a resource that has value
- Publisher of data as open data - they strive to educate and equip their people to make confident decisions about what they can and can't share and open, and manage their private and sensitive data rather than avoid it altogether.

A number of government open data publishers have also piloted the use of Open Data Certificates. A localised Australian version of open data certificates will be available soon. Open data certificates are a great way to help educate publishers of open data, and help to build trust in using open data by providing a mark of confidence for users of open data.

Transport agencies adopting GTFS standard to enable citizens to make informed decisions about their travel on public transport as well as the Bureau of Meteorology has been effective in this area.

Questions on data linkage

Which datasets, if linked or coordinated across public sector agencies, would be of high value to the community, and how would they be used?

Some examples include:

- Transport network - enable cross-jurisdictional transactions such as heavy vehicle permitting. http://www.anzlic.gov.au/fsdf-themes-datasets
- Health, emergency response, and social services - which happen at every level of government
- Cross agency linking around critical services

Which rules, regulations or policies create unnecessary or excessive barriers to linking datasets?

The PSMA business model might be suggested as an example as representing challenge in this area.

How can Australia's government agencies improve their sharing and linking of public

sector data? What lessons or examples from overseas should be considered?

The key to improvement in the UK for example has been access to tools and education offered by the ODI.

2. **Private sector data**
   Questions on high value private sector data

   What private sector datasets should be considered high-value data to: public policy; researchers and academics; other private sector entities; or the broader community?

   Examples include:
   - Enrolment data at tertiary education facilities
   - Data by NFPs, NGOs and private organisations who are delivering services which Government used to do but has outsourced
   - Private transport operator data
   - Flood studies, flood models (+underlying raw data)
   - Environmental impact assessments (+underlying raw data)
   - Pricing data (eg airline tickets, health insurance products, telecommunications products and services)

   In each case cited, what characteristics define such datasets? What would be the public policy rationale for any associated government intervention?

   Examples of characteristics would include standard descriptors and standard definitions.

   What benefits would the community derive from increasing the availability and use of private sector data?

   The principal benefit is better informed decision-making.

   Questions on access to private sector data

   Are there any legislative or other impediments that may be unnecessarily restricting the availability and use of private sector data? Should these impediments be reduced or removed?

   We have not yet undertaken any formal consultation in relation to this, however in our experience so far it would seem it is predominantly a cultural resistance. There is also a lack of compelling examples of benefits of using private sector data to motivate. The cost of sharing data is considered to outweigh any benefit and there is a fear of disclosing data will reduce competitive advantage.

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

We have not yet undertaken any formal consultation in relation to this, however in our experience so far there is a fear of consequences of combining data sets together to create unintended consequences, a fear of mismanaging private or sensitive data, and a fear of losing competitive advantage by giving others access to their data.

What principles, protocols or legislative requirements could manage the concerns of private sector data owners about increasing the availability of their data?

Security and knowledge about who / which users can access and use the data and for what purposes, and effective management of access and use permissions.

Should the collection, sharing and release of private sector data be standardised in some way? How could this be done and what would be the benefits and costs? What would standards that are 'fit for purpose' look like?

In our experience so far, it would seem that the private sector is reasonably mature in terms of appreciating the value of data and how it can be used to inform decision making. Standards in terms of descriptors and definitions would be helpful.

To what extent can voluntary data sharing arrangements — between businesses / between businesses and consumers / involving third party intermediaries — improve outcomes for the availability and use of private data? How could participation levels be increased?

The recently announced open data partnership between the Queensland Government and ODI Queensland includes the development of an education and training program to help SMEs learn how to discover and access open data, to use and consume it to provide new knowledge and insight for their own use. We would suggest the involvement of trusted third party intermediaries in this area is critical.

Would such voluntary arrangements raise competition issues? How might this change if private sector information sharing were mandated? Is authorisation (under the Competition and Consumer Act 2010 (Cth)) relevant?

We are not in a position to assist with this question at this stage.

What role can governments usefully play in promoting the wider availability of private datasets that have the potential to deliver substantial spillover benefits?

In our experience most private sector organisations are open to the concept of sharing their data for the social good, however they are less enthusiastic about others profiting from their data. Promotion of use case examples which can demonstrate value and provide some ‘what’s in it for me’ motivation would be helpful. The role of government might be to work with industry to create standards for descriptors and definitions for example.

How can the sharing and linking of private sector data be improved in Australia? What lessons or examples from overseas should be considered?

The ODI has published some case studies in the UK that may be of some assistance.