Submission from Ms. Bernadette Hyland-Wood

Affiliation.   The University of Queensland

Email.   ███████████████████████

Declared Interests / affiliations: Former international working group co-chair Government Linked Data Working Group, W3C (2011-2013), Open Data Consultant to the U.S. Environmental Protection Agency (EPA) 2011-2017 and UN FAO, U.S. Library of Congress and Government Publishing Office (GPO) 2009-2017, former small business member of the Washington DC-based Data Coalition.

1 Do you think the distinction between data sharing and data release is clear? How could this distinction be clearer?

ONDC has defined the terms and how it uses them, however that usage is not generally used in the scholarly literature on government information, nor within the broader open government data community. It is more common to read definitions along the lines of:

Public sector data or *government data* is digital data collected, used or held by government in fulfilment of its various functions. This data is vast and includes weather data to administrative data on government operations. *Open* in the context of data sharing means non-proprietary, made available and accessible on the public Web.

Open government data is that subset of public sector or government data published by on the public Web, typically on publicly accessible government websites, for use by anyone, anywhere, without license restrictions, and should not contain security attributes, or personally-identifiable information (W3C, 2013; Kalampokis, Tambouris, & Tarabanis, 2011).

The terms 'open government data' and 'public data', are often used interchangeably in the literature.

As for "to share", it is not surprising that ONDC may be encountering confusion on what this means to the many stakeholders. (I discovered the same thing after two years of PhD research, following 15+ years of industry experience working on data sharing, really.)

Submission by B Hyland-Wood to the Australian Office of the National Data Commissioner (ONDC) on Data Sharing and Release Legislative Reforms

Submission date: 15-Oct 2019   **1**

There are many interpretations, however I offer the following table that can be adapted for the ONDC purposes, should it prove useful to tease out the various forms of 'sharing'.

Case	Use case description for government data (closed data)
1	Not at all, i.e., siloes with sensitive data for law enforcement or military purposes
2	Internally within a team or department, i.e., spreadsheets, documents, presentations
3	Internally with other trusted groups, i.e., health services data with another government agency
4	Internally via a cross-governmental security regime
5	Externally with designated allies, i.e., Five Eyes alliance (FVEY)
6	Externally with contractors with clearances

Case	Use case description (variations of open government data)
7	Externally with trusted network, e.g., CSIRO scientists collaborating, OpenPHACTs life sciences community. Requires membership, non-sensitive data, "pre-competitive data"
8	Externally with partnership agreement, e.g., GeoScience Australia and mining industry, third party aggregators
9	Externally via Freedom of Information Act (FOIA) request
10	Externally if you know where to look for public data, often has an application programming interface (API). May require pre-authorization and service agreement
11	Externally on a public website, typically with a .gov or similar, domain. Includes visualizations, dashboards on public spending, etc.

Note each of the above 11 use cases involves "sharing", it is a matter of with whom. Therefore, the currently proposed ONDC language where sharing is internally is likely to cause confusion with the larger, international data sharing community that has a decade, and arguably three decades, depending upon where you delineate the start.

Recommendation:

Be explicit about with whom data is being shared, rather than trying to boil it down to the proposed terminology of "Shared Data" and "Open Release Data" which is just too simplistic.

While it is important to keep messaging clear, there also needs to be realistic assessment of the use cases that simply don't fit the proposed "shared" versus "open" model.

2 What are the challenges for open release of public sector data?

Submission by B Hyland-Wood to the Australian Office of the National Data Commissioner (ONDC) on Data Sharing and Release Legislative Reforms

Submission date: 15-Oct 2019   2

The factors hindering data sharing are extensively detailed in the government information literature. Challenges include: (a) deficient data management practices and policies; (b) lack of a supportive culture, incentives and training; (c) risk avoidance; (d) privacy, security and legal barriers; and (e) politicization of public data.

Technical or procedural improvements alone are insufficient (Dawes, 2012; Huijboom & Van den Broek, 2011; Zuiderwijk & Janssen, 2014), and ultimately it depends on the willingness of individuals, often working in collaboration with others to improve how public data is prioritized and shared.

Interviews with data champions and data custodians point to the importance of moving from deficient data management practices to a data sharing culture.

Again, the subject of a PhD thesis (forthcoming).

Having advocated (for over a decade), for increased availability of publicly-funded non-sensitive data to enhance scientific research (at the international level), it comes down to the people involved, their willingness to genuinely engage in the current international community of open government and open government data experts.

While the goals and objectives (access to data policy and technology experts, budget, etc), of each country are different, there is a tremendous amount we can learn by working together. This not 'motherhood & apple pie', but a genuine insight that the countries that work together are more effective, and those who 'go it alone' re-create the wheel and end up with hard to maintain siloes, and lack of community engagement.

Recommendation:

There has been a considerable amount of well-informed data science that has gone into creating Web data standards, for example, and Australia need not recreate the wheel, but engage, leverage and extend what is there.

The ONDC call also help generate some 'pull' for data science expertise that will stimulate supply of university students enrolled in data science and cyber security programs. It would be nice if they had some place other than foreign governments, Amazon and Google to join upon graduation from Australian tertiary institutions.

3 Do you think the Data Sharing and Release legislative framework will achieve more streamlined and safer data sharing?

The resulting DS&R legislation will depend on how it is informed, who drafts it and the negotiations the ultimately result in passed legislation.

Advice, keep up the open, consultative discussion to get the best outcomes. It isn't an end point that ONDC will arrive at – it is process. Small is beautiful. Don't try to boil the ocean and

Submission by B Hyland-Wood to the Australian Office of the National Data Commissioner (ONDC) on Data Sharing and Release Legislative Reforms

Submission date: 15-Oct 2019 **3**

attempt to address this massive area in one go. Break it apart into low hanging fruit and address the scoped deliverables that you appear to understand clearly.

Having worked in Washington DC during the period that the US DATA Act was prepared and passed (2013-2014), see https://www.congress.gov/bill/113th-congress/senate-bill/994 the benefits of having a 'large round table' to gather the data science experts, W3C rep, industry, policy advisors, and even several Congressional members, resulted in bipartisan accomplishments.

The DATA Act was successful because the data lobby (Data Transparency Coalition with Hudson Hollister, Exec Director), maintained a very "big tent" and encouraged SMEs and large end of town to be heard. There was significant consultation and engagement with policy advisors, and together, they climbed a steep learning curve and ultimately prevailed in an astonishing 5 years (from concept and a one-man band to signed US Legislation). It was the only bipartisan bill that passed in the timeframe, I believe. And, the DATA Act was all about financial spending by government which nearly all legislators wanted increased visibility on. That is a textbook case of what worked well.

<u>**Recommendation**</u>:

Put together a 'working group', 'council', (whatever you want to call it) that is diverse and inclusive in its composition (meaning not only public service people with economics backgrounds :-), and also include people with serious technical and policy chops who are part of centres of excellence in data science, cyber security, AI, healthcare/life sciences research, data policy, Indigenous data sovereignty, open government data forums and partnerships. A group that is capped at around 15-20 ideally. The members must have a clear charter, deliverables and meet from time to time face-to-face. Diversity of opinions, perspectives and representation really important, just to make sure that is clear.

Here are a few groups that should be on your radar to be engaging with:

Open Government Partnership https://www.opengovpartnership.org/ ,

Open Data Charter, https://opendatacharter.net/

The Australian Government Linked Data Working Group, http://www.linked.data.gov.au/, operated via an MOU between about six Australian agencies.

4 What do you think about the name, Data Sharing and Release Act?

It is fine. No issues.

Submission by B Hyland-Wood to the Australian Office of the National Data Commissioner (ONDC) on Data Sharing and Release Legislative Reforms

Submission date: 15-Oct 2019   **4**

5 Do the purposes for sharing data meet your expectations? What about precluded purposes?

It is in keeping with the Government's responsibilities to focus on improving data sharing to improve availability of high quality, authoritative data to inform public policy, support publicly-funded R&D, and improve government service delivery.

Yes, that should remain the focus.

6 What are your expectations for commercial uses? Do we need to preclude a purpose, or do the Data Sharing Principles and existing legislative protections work?

Several OECD countries have a long history of collecting and sharing publicly funded data that supports applied research and industry globally. Sharing non-sensitive public data (in the broad sense) with trusted partners, industry, NGOs, is both realistic and delivers economic benefits to the nation and internationally. It's also a valuable form of soft power.

For example, Australia is the beneficiary of tens of billions of dollars (USD) worth of taxpayer funded open data (Global Positioning System) or GPS data collected by the US Air Force, that is made available for peaceful, civilian purposes. Agriculture, logistics and university research centres, (not to mention the diverse range of food delivery services popping up in Australian CBDs) would grind to a halt if that open data were not freely shared by the US Government's GPS program.

Note: I'm reading between the lines as to why a $150M commitment to work on space R&D with NASA was recently announced by Minister Morrison after visiting the USA. It would be great to see the Australian space agency get their own satellites (cubesat) infrastructure operational (Australia has some good specialist rocket science expertise, AI, blockchain, etc.). That is tangentially related to data sharing legislation.

The public should not be scared of supporting commercial use – within reason. Geoscience Australia produces a phenomenal amount of open data that is used by both government and the mining industry. Billions of dollars of revenue are attributable to that investment by the Australian Government. Same for CSIRO's data collection on Earth observations.

If you haven't heard CSIRO's Dr. Adrian Turners talk on ABC Radio's Science Show, titled "Australia at the back of the pack in digital innovation", have a listen! https://www.abc.net.au/radionational/programs/scienceshow/anyone-fancy-$315-billion/11326570

7 Do you think the Data Sharing Principles acknowledge and treat risks appropriately? When could they fall short?

Submission by B Hyland-Wood to the Australian Office of the National Data Commissioner (ONDC) on Data Sharing and Release Legislative Reforms

Submission date: 15-Oct 2019   5

Entirely depends on the details and implementation. Learn from the experts, nationally and internationally. Get them to advise on de-anonymization, or at a minimum, read the literature. See Anonymizing Health Data, by Dr. Luk Arbuckle and Prof Khaled El Emam https://www.oreilly.com/library/view/anonymizing-health-data/9781449363062/

This O'Reilly link is dated (2011) but gives a reasonable 'birds eye view', see https://www.oreilly.com/ideas/anonymize-data-limits

Same with end-to-end encryption debate. Providing "backdoors" for legitimate law enforcement leaves the backdoor for anyone, good or bad actor. The premise is naïve. The ONDC could do some real good facilitating experts to meet & greet policy makers and ministers so they don't embarrass themselves on the international stage. Misguided or misinformed policy on information sharing platforms do not keep citizen safe.

8 Is the Best Practice Guide to Applying Data Sharing Principles helpful? Are there areas where the guidance could be improved?

Yes, very useful. It is the maximum size that people will typically read. I've cited it in my research ☺

Recommendation:

One idea may be break it into smaller discrete components with high level info on 'governance and data sharing agreements' and then allow a user to download / share that info with others, without having to wade through 50+ pages. Organise thematically or by person using it, e.g., data custodian/Web developer, data champion/manager, procurement personnel, etc.

Glossary of terms needs some additional iteration, incomplete given all that ONDC and DS&R will encompass. Clear, concise terms and definitions are very useful and will be cited if you do a good job.

Please make this a web page with anchors, so people can cite in papers, websites, etc. Example, see W3C linked data glossary https://www.w3.org/TR/ld-glossary/

9 Do the safeguards address key privacy risks?

Not my field, unable to comment.

10 Are the core principles guiding the development of accreditation criteria comprehensive? How else could we improve and make them fit for the future?

Did not review. Overall, I think it is a good idea as long as it is voluntary, like a good housekeeping seal of approval. If the process and/or training is high quality, it may be very helpful in training & 'aculturating' (StatsNZ Paul Stone's word) the public service to a data sharing culture. Also, DIIS is doing some great work on shifting the data culture.

Submission by B Hyland-Wood to the Australian Office of the National Data Commissioner (ONDC) on Data Sharing and Release Legislative Reforms

Submission date: 15-Oct 2019   6

Leverage in-government expertise, there are some extraordinary people currently doing great work. ONDC can extend reach by working with those already in the data community.

11 Are there adequate transparency and accountability mechanisms built into the framework, including Data Sharing Agreements, public registers and National Data Commissioner review and reporting requirements?

I think the ONDC has done an excellent job communicating the progress of each iteration in a thoughtful, reflective and transparent manner. You've made workshops open and inclusive, and genuinely appear to be seeking and synthesising the voices in a very complicated and nuanced landscape. I attended two workshops (Canberra and Brisbane) and found them informative, well conducted and with a excellent engagement by the public, uni and industry.

ONDC gets 10/10 of facilitation and communication.

12 Have we achieved the right balance between complaints, redress options and review rights?

Unable to evaluate.

13 Have we got our approach to enforcement and penalties right for when things go wrong? Will it deter non-compliance while encouraging greater data sharing?

I haven't been able to review the proposed approach.

I will comment that there must be very clearly articulated penalties and enforcement on the removal or obfuscation of publicly-funded data.

The practical relevance of government data sharing has increased with the rise of the current 'populist zeitgeist' (Mudde, 2004) as research findings and regulatory data deemed inconsistent with a new administration's agenda, is removed or obscured (Beeler, 2017; Brady, 2016; Center for Science and Democracy, 2018; Kang & Shear, 2017). Data is ephemeral and can be deleted with literally a few strokes from a keyboard, issuing a command to a remote cloud-based server. The loss of data from servers, scholarly citations, dependent services, and ultimately institutional memory is a complex and nuanced area.

Happy to discuss this further as I have firsthand experience of removal (through de-funding), of the largest Web service providing 30+ years of US pollution emissions, toxic chemicals and regulated facilities (US EPA open data site). The loss / removal of previously open data, especially those who work in alliances and partnerships, affects other nations. This should be considered in the ONDC DS&R framework – the issue of dependence, linking, ongoing support & maintenance.

14 What types of guidance and ongoing support from the National Data Commissioner will provide assurance and enable safe sharing of data?

Submission by B Hyland-Wood to the Australian Office of the National Data Commissioner (ONDC) on Data Sharing and Release Legislative Reforms

Submission date: 15-Oct 2019   7

Engagement with experts in data sciences, social sciences on topics including Indigenous Data Sovereignty and data ethics, will go a long way.

Don't recreate the wheel, there has been a lot of excellent work done and available for the Australian Government to extend and leverage, and make your own.

Recommended resources:

ACOLA Horizon Scanning Report: The Effective and Ethical Development of Artificial Intelligence: An opportunity to improve our wellbeing

https://acola.org/wp-content/uploads/2019/07/hs4_artificial-intelligence-report.pdf

^^ Chapters 6 and 7 are dedicated to "data" and "inclusive data", respectively.

Also, see "Good Data", (ed) Prof Angela Daly (Chinese University & QUT), Dr. S.Kate Devitt (Australian Defence Science & Technology Group), and Dr. Monique Mann (QUT)

https://networkcultures.org/blog/publication/tod-29-good-data/

^^ Provides important diversity of voices, and concrete steps on how we can start realizing good data practice.

Submission by B Hyland-Wood to the Australian Office of the National Data Commissioner (ONDC) on
Data Sharing and Release Legislative Reforms

Submission date: 15-Oct 2019   **8**