Personal submission to the Productivity Commission Review on Public Sector Data

My name is Pia Waugh and this is my personal submission to the [Productivity Commission Review on Public Sector Data](#). It does not reflect the priorities or agenda of my employers past, present or future, though it does draw on my expertise and experience in driving the open data agenda and running data portals in the ACT and Commonwealth Governments from 2011 till 2015. I was invited by the Productivity Commission to do a submission and thought I could provide some useful ideas for consideration. I note I have been on maternity leave since January 2016 and am not employed by the Department of Prime Minister and Cabinet or working on data.gov.au at the time of writing this submission. This submission is also influenced by my work and collaboration with other Government jurisdictions across Australia, overseas and various organisations in the private and community sectors. I'm more than happy to discuss these ideas or others if useful to the Productivity Commission.

I would like to thank all those program and policy managers, civic hackers, experts, advocates, data publishers, data users, public servants and vendors whom I have had the pleasure to work with and have contributed to my understanding of this space. I'd also like to say a very special thank you to the Australian Government Chief Technology Officer, John Sheridan, who gave me the freedom to do what was needed with data.gov.au, and to Allan Barger who was my right hand man in rebooting the agenda in 2013, supporting agencies and helping establish a culture of data publishing and sharing across the public sector. I think we achieved a lot in only a few years with a very small but highly skilled team. A big thank you also to Alex Sadleir and Steven De Costa who were great to work with and made it easy to have an agile and responsive approach to building the foundation for an important piece of data infrastructure for the Australian Government.

Finally, this is a collection of some of my ideas and feedback for use by the Productivity Commission however, it doesn't include everything I could possibly have to say on this topic because, frankly, we have a small baby who is taking most of my time at the moment. Please feel free to add your comments, criticisms or other ideas to the comments below! It is all licensed as *Creative Commons 4.0 By Attribution*, so I hope it is useful to others working in this space.

Table of Contents

The Importance of Vision

Without a vision, we stumble blindly in the darkness. Without a vision, the work and behaviours of people and organisations are inevitably driven by other competing and often short term priorities. In the case of large and complex organisms like the Australian Public Service, if there is no cohesive vision, no clear goal to aim for, then each individual department is going to do things their own way, driven by their own priorities, budgets, Ministerial whims and you end up with what we largely have today: a cacophony of increasingly divergent approaches driven by tribalism that make collaboration, interoperability, common systems and data reuse impossible (or prohibitively expensive).

If however, you can establish a common vision, then even a strongly decentralised system can converge on the goal. If we can establish a common vision for public data, then the implementation of data programs and policies across the APS should become naturally more consistent and common in practice, with people naturally motivated to collaborate, to share expertise, and to reuse systems, standards and approaches in pursuit of the same goal.

My vision for public data is two-pronged and a bit of a paradigm shift: **data by design** and **gov as an API!** "Data by design" is about taking a data driven approach to the business of government and "gov as an API" is about changing the way we use, consume, publish and share data to properly enable a data driven public service and a broader network of innovation. The implementation of these ideas would create *mashable government* that could span departments, jurisdictions and international boundaries. In a heavily globalised world, no government is in isolation and it is only by making government data, content and services API enabled and reusable/interfacable, that we, collectively, can start to build the kind of analysis, products and services that meet the necessarily cross jurisdictional needs of all Australians, of all people.

More specifically, my vision is a data driven approach to the entire business of government that supports:

- evidence based and iterative policy making and implementation;
- transparent, accountable and responsible Government;
- an open competitive marketplace built on mashable government data, content and services; and
- a more efficient, effective and responsive public service.

What this requires is not so simple, but is utterly achievable if we could embed a more holistic whole of government approach in the work of individual departments, and then identify and fill the gaps through a central capacity that is responsible for driving a whole of government approach. Too often we see the data agenda oversimplified into what outcomes are desired (data visualisations, dashboards, analysis, etc) however, it is only in establishing multipurpose *data infrastructure* which can be reused for many different purposes that we will enable the kind of insights, innovation, efficiencies and effectiveness that all the latest reports on realising the value of data allude to. Without actual data, all the reports, policies, mission statements, programs and governance committees are essentially wasting time. But to get better government data, we need to build *capacity and motivation* in the public sector. We need to build a *data driven culture* in government. We also need to *grow consumer confidence* because a)

demand helps drive supply, and b) if data users outside the public sector don't trust that they can find, use and rely upon at least some government data, then we won't ever see serious reuse of government data by the private sector, researchers, non-profits, citizens or the broader community.

Below is a quick breakdown of each of these priorities, followed by specific recommendations for each:

data infrastructure that supports multiple types of reuse. Ideally all data infrastructure developed by all government entities should be built in a modular, API enabled way to support data reuse beyond the original purpose to enable greater sharing, analysis, aggregation (where required) and publishing. It is often hard for agencies to know what common infrastructure already exists and it is easy for gaps to emerge, so another part of this is to map the data infrastructure requirements for all government data purposes, identify where solutions exist and any gaps. Where whole of government approaches are identified, ***common data infrastructure*** should be made available for whole of government use, to reduce the barrier to publishing and sharing data for departments. Too often, large expensive data projects are implemented in individual agencies as single purpose analytics solutions that don't make the underlying data accessible for any other purpose. If such projects separated the data infrastructure from the analytics solutions, then the data infrastructure could be built to support myriad reuse including multiple analytics solutions, aggregation, sharing and publishing. If government data infrastructure was built like any other national infrastructure, it should enable a competitive marketplace of analysis, products and service delivery both domestically and globally. A useful analogy to consider is the example of roads. Roads are not typically built just from one address to another and are certainly not made to only support certain types of vehicles. It would be extremely inefficient if everyone built their own custom roads and then had to build custom vehicles for each type of road. It is more efficient to build common roads to a minimum technical standard that any type of vehicle can use to support both immediate transport needs, but also unknown transport needs into the future. Similarly we need to build multipurpose data infrastructure to support many types of uses.

greater publisher capacity and motivation to share and publish data. Currently the range of publishing capacity across the APS is extremely broad, from agencies that do nothing to agencies that are prolific publishers. This is driven primarily by different cultures and responsibilities of agencies and if we are to improve the use of data, we need to improve the supply of data across the entire public sector. This means education and support for agencies to help them understand the value to their BAU work. The time and money saved by publishing data, opportunities to improve data quality, the innovation opportunities and the ability to improve decision making are all great motivations once understood, but generally the data agenda is only pitched in political terms that have little to no meaning to data publishers. Otherwise there is no natural motivation to publish or share data, and the strongest policy or regulation in the world does not create sustainable change or effective outcomes if you cannot establish a motivation to comply. Whilst ever publishing data is seen as merely a compliance issue, it will be unlikely for agencies to invest the time and skills to publish data well, that is, to publish the sort of data that consumers want to use.

greater consumer confidence to improve the value realised from government data. Supply is nothing without demand and currently there is a relatively small (but growing) demand for government data, largely because people won't use what they don't trust. In the current landscape is difficult to find data and even if one can find it, it is often not machine readable or not freely available, is out of date and generally hard to use. There is not a high level of consumer confidence in what is provided by government so many people don't even bother to look. If they do look, they find myriad data sources of ranging quality and inevitably waste many hours trying to get an outcome. There is a reasonable demand for data for research and the research community tends to jump through hoops - albeit reluctantly and at great cost - to gain access to government data. However, the private and civic sectors are yet to seriously engage apart form a few interesting outliers. We need to make finding and using *useful* data easy, and start to build consumer confidence or we will never even scratch the surface of the billions of dollars of untapped potential predicted by various studies. The ***data infrastructure*** section is obviously an important part of building consumer confidence as it should make it easier for consumers to find and have confidence in what they need, but it also requires improving the data culture across the APS, better outreach and communications, better education for public servants and citizens on how to engage in the agenda, and targeted programs to improve the publishing of data already in demand. What we don't need is yet another "tell us what data you want" because people want to see progress.

a data driven culture that embeds in all public servants an understanding of the role of data in the every day work of the public service, from program management, policy development, regulation and even basic reporting. It is important to take data from being seen as a specialist niche delegated only to highly specialised teams and put data front and centre as part of the responsibilities of all public servants - especially management - in their BAU activities. Developing this culture requires education, data driven requirements for new programs and policies, some basic skills development but mostly the proliferation of an awareness of what data is, why it is important, and how to engage appropriate data skills in the BAU work to ensure a data driven approach. Only with data can a truly evidence driven approach to policy be taken, and only with data can a meaningful iterative approach be taken over time.

Finally, obviously the approach above requires an appropriately skilled team to drive policy, coordination and implementation of the agenda in collaboration with the broader APS. This team should reside in a central agenda to have whole of government imprimatur, and needs a mix of policy, commercial, engagement and technical data skills. The experience of data programs around the world shows that when you split policy and implementation, you inevitably get both a policy team lacking in the expertise to drive meaningful policy and an implementation team paralysed by policy indecision and an unclear mandate. This space is changing so rapidly that policy and implementation need to be agile and mutually reinforcing with a strong focus on getting things done.

As we examine the interesting opportunities presented by new developments such as blockchain and big data, we need to seriously understand the shift in paradigm from scarcity to surplus, from centralised to distributed systems, and from pre-planned to iterative approaches, if we are to create an effective public service for the 21st century.

There is already a lot of good work happening, so the recommendations in this submission are meant to improve and augment the landscape, not replicate. I will leave areas of specialisation to the specialists, and have tried to make recommendations that are supportive of a holistic approach to developing a data-driven public service in Australia.

Current Landscape

There has been progress in recent years towards a more data driven public sector however, these initiatives tend to be done by individual teams in isolation from the broader public service. Although we have seen some excellent exemplars of big data and open data, and some good work to clarify and communicate the intent of a data driven public service through policy and reviews, most projects have simply expanded upon the status quo thinking of government as a series of heavily fortified castles that take the extraordinary effort of letting in outsiders (including other departments) only under strictly controlled conditions and with great reluctance and cost. There is very little sharing at the implementation level (though an increasing amount of sharing of ideas and experience) and very rarely are new initiatives consulted across the APS for a whole of government perspective. Very rarely are actual data and infrastructure experts encouraged or supported to work directly together across agency or jurisdiction lines, which is a great pity. Although we have seen the idea of the value of data start to be realised and prioritised, we still see the implementation of data projects largely delegated to small, overworked and highly specialised internal teams that are largely not in the habit of collaborating externally and thus there is a lot of reinvention and diversity in what is done.

If we are to realise the real benefits of data in government and the broader economy, we need to challenge some of the status quo thinking and approaches towards data. We need to consider government (and the data it collects) as a platform for others to build upon rather than the delivery mechanism for all things to all people. We also need to better map what is needed for a data-driven public service rather than falling victim to the attractive (and common, and cheap) notion of simply identifying existing programs of work and claiming them to be sufficient to meet the goals of the agenda.

Globally this is still a fairly new space. Certain data specialisations have matured in government (eg. census/statistics, some spatial, some science data) but there is still a lack of a cohesive approach to data in any one agency. Even specialist data agencies tend to not look beyond the specialised data to have a holistic data driven approach to everything. In this way, it is critical to develop a holistic approach to data at all levels of the public service to embed the principles of data driven decision making in everything we do. Catalogues are not enough. Specialist data projects are not enough. Publishing data isn't enough. Reporting number of datasets quickly becomes meaningless. We need to measure our success in this space by how well data is helping the public service to make better decisions, build better services, develop and iterate responsive and evidence based policy agendas, measure progress and understand the environment in which we operate.

Ideally, government agencies need to adopt a dramatic shift in thinking to assume in the first instance that the best results will be discovered through collaboration, through sharing, through helping people help themselves. There also needs in the APS to be a shift away from thinking that a policy, framework, governance structure or other artificial constructs are sufficient outcomes. Such mechanisms can be useful, but they can also be

a distraction from getting anything tangible done. Such mechanisms often add layers of complexity and cost to what they purport to achieve. Ultimately, it is only what is actually implemented that will drive an outcome and I strongly believe an outcomes driven approach must be applied to the public data agenda for it to achieve its potential.

References

In recent years there has been a lot of progress. Below is a quick list to ensure they are known and built upon for the future. It is also useful to recognise the good work of the government agencies to date.

- Public Data Toolkit - the data.gov.au team have pulled together a large repository of information, guidance and reports over the past 3 years on our open data toolkit at http://toolkit.data.gov.au. There are also some useful contributions from the Department of Communications Spatial Policy Branch. The Toolkit has links to various guidance from different authoritative agencies across the APS as well as general information about data management and publishing which would be useful to this review.
- The Productivity Commission is already aware of the Legislative and other Barriers Workshop I ran at PM&C before going on maternity leave, and I commend the outcomes of that session to the Review.
- The Financial Sector Inquiry (the "Murray Inquiry") has some excellent recommendations regarding the use of data-drive approaches to streamline the work and reporting of the public sector which, if implemented, would generate cost and time savings as well as the useful side effect of putting in place data driven practices and approaches which can be further leveraged for other purposes.
- Gov 2.0 Report and the Ahead of the Game Report – these are hard to find copies of online now, but have some good recommendations and ideas about a more data centric and evidence based public sector and I commend them both to the Review. I'm happy to provide copies if required.
- There are many notable APS agency efforts which I recommend the Productivity Commission engage with, if they haven't already. Below are a few I have come across to date, and it is far from an exhaustive list:
    - PM&C (Public Data Management Report/Implementation & Public Data Policy Statement)
    - Finance (running and rebooting data.gov.au, budget publishing, data integration in GovCMS)
    - ABS (multi agency arrangement, ABS.Stat)
    - DHS (analytics skills program, data infrastructure and analysis work)
    - Immigration (analytics and data publishing)
    - Social Services (benefits of data publishing)
    - Treasury (Budget work)
    - ANDS (catalogue work and upskilling in research sector)
    - NDI (super computer functionality for science)
    - ATO (smarter data program, automated and publications data publishing, service analytics, analytics, dev lab, innovationspace)
    - Industry (Lighthouse data integration and analysis, energy ratings data and app)
    - CrimTRAC and AUSTRAC (data collection, consolidation, analysis, sharing)
- Other jurisdictions in Australia have done excellent work as well and you can see a list (hopefully up to date) of portals and policies on the Public Data Toolkit. I

recommend the Productivity Commission engage with the various data teams for their experiences and expertise in this matter. There are outstanding efforts in all the State and Territory Governments involved as well as many Local Councils with instructive success stories, excellent approaches to policy, implementation and agency engagement/skills and private sector engagement projects.

Current risks/issues

There are a number of issues and risks that exist in pursuing the current approach to data in the APS. Below are some considerations to take into account with any new policies or agendas to be developed.

- There is significant duplication of infrastructure and investment from building bespoke analytics solutions rather than reusable data infrastructure that could support multiple analytics solutions. Agencies build multiple bespoke analytics projects without making the underpinning data available for other purposes resulting in duplicated efforts and under-utilised data across government.
- Too much focus on pretty user interfaces without enough significant investment or focus on data delivery.
- Discovery versus reuse – too many example of catalogues linking to dead data. Without the data, discovery is less than useful.
- Limitations of tech in agencies by ICT Department – often the ICT Department in an agency is reticent to expand the standard operating environment beyond the status quo, creating an issue of limitation of tools and new technologies.
- Copyright and legislation - particularly old interpretations of each and other excuses to not share.
- Blockers to agencies publishing data (skills, resources, time, legislation, tech, competing priorities e.g. assumed to be only specialists that can do data).
- Often activities in the public sector are designed to maintain the status quo (budgets, responsibilities, staff count) and there is very little motivation to do things more efficiently or effectively. We need to establish these motivations for any chance to be sustainable.
- Public perceptions about the roles and responsibilities of government change over time and it is important to stay engaged when governments want to try something new that the public might be uncertain about. There has been a lot of media attention about how data is used by government with concerns aired about privacy. Australians are concerned about what Government plans to do with their data. Broadly the Government needs to understand and engage with the public about what data it holds and how it is used. There needs to be trust built to both improve the benefits from data and to ensure citizen privacy and rights are protected. Where government wants to use data in new ways, it needs to prosecute the case with the public and ensure there are appropriate limitations to use in place to avoid misuse of the data. Generally, where Australians can directly view the benefit of their data being used and where appropriate limitations are in place, they will probably react positively. For example, tax submission are easier now that their data auto-fills from their employers and health providers when completing Online Tax. People appreciate the concept of having to only update their details once with government.

Benefits

I agree with the benefits identified by the Productivity Commission discussion paper however I would add the following:

- Publishing government data, if published well, enables a competitive marketplace of service and product delivery, the ability to better leverage public and academic analysis for government use and more broadly, taps into the natural motivation of the entire community to innovate, solve problems and improve life.
- Establishing authoritative data – often government is the authoritative source of information it naturally collects as part of the function of government. When this data is not then publicly available (through anonymised APIs if necessary) then people will use whatever data they can get access to, reducing the authority of the data collected by Government
- A data-drive approach to collecting, sharing and publishing data enables true iterative approaches to policy and services. Without data, any changes to policy are difficult to justify and impossible to track the impact, so data provides a means to support change and to identify what is working quickly. Such feedback loops enable iterative improvements to policies and programs that can respond to the changing financial and social environment the operate in.
- Publishing information in a data driven way can dramatically streamline reporting, government processes and decision making, freeing up resources that can be used for more high value purposes.

Public Sector Data Principles

The Public Data Statement provides a good basis of principles for this agenda. Below are some principles I think are useful to highlight with a brief explanation of each.

Principles:
- *build for the future* - legacy systems will always be harder to deal with so agencies need to draw a line in the sand and ensure new systems are designed with data principles, future reuse and this policy agenda in mind. Otherwise we will continue to build legacy systems into the future. Meanwhile, just because a legacy system doesn't natively support APIs or improved access doesn't mean you can't affordably build middleware solutions to extract, transform, share and publish data in an automated way.
- *data first* - wherever data is used to achieve an outcome, publish the data along with the outcome. This will improve public confidence in government outcomes and will also enable greater reuse of government data. For example, where graphs or analysis are published also publish the data. Where a mobile app is using data, publish the data API. Where a dashboard is set up, also provide access to the underpinning data.
- *use existing data, from the source where possible* - this may involve engaging with or even paying for data from private sector or NGOs, negotiating with other jurisdictions or simply working with other government entities to gain access.
- *build reusable data infrastructure first* - wherever data is part of a solution, the data should be accessible through APIs so that other outcomes and uses can be realised, even if the APIs are only used for internal access in the first instance.
- *data driven decision making* to support iterative and responsive policy and implementations approaches - all decisions should be evidence based, all projects,

policies and programs should have useful data indicators identified to measure and monitor the initiative and enable iterative changes backed by evidence.

- *consume your own data and APIs* - agencies should consider how they can better use their own data assets and build access models for their own use that can be used publicly where possible. In consuming their own data and APIs, there is a better chance the data and APIs will be designed and maintained to support reliable reuse. This could raw or aggregate data APIs for analytics, dashboards, mobile apps, websites, publications, data visualisations or any other purpose.
- *developer empathy* - if government agencies start to prioritise the needs of data users when publishing data, there is a far greater likelihood the data will be published in a way developers can use. For instance, no developer likes to use PDFs, so why would an agency publish data in a PDF (hint: there is no valid reason. PDF does not make your data more secure!).
- *standardise where beneficial but don't allow the perfect to delay the good* - often the focus on data jumps straight to standards and then multi year/decade standards initiatives are stood up which creates huge delays to accessing actual data. If data is machine readable, it can often be used and mapped to some degree which is useful, more useful than having access to nothing.
- *automate, automate, automate!* - where human effort is required, tasks will always be inefficient and prone to error. Data collection, sharing and publishing should be automated where possible. For example, when data is regularly requested, agencies should automate the publishing of data and updates which both reduces the work for the agency and improves the quality for data users.
- *common platforms* - where possible agencies should use existing common platforms to share and publish data. Where they need to develop new infrastructure, efforts should be made to identify where new platforms might be useful in a whole of government or multi agency context and built to be shared. This will support greater reuse of infrastructure as well as data.
- *a little less conversation a little more action* – the public service needs to shift from talking about data to doing more in this space. Pilot projects, experimentation, collaboration between implementation teams and practitioners, and generally a greater focus on getting things done.

Recommendations for the Public Data agenda

Strategic

1. Strong Recommendation: Develop a holistic vision and strategy for a data-driven APS. This could perhaps be part of a broader digital or ICT strategy, but there needs to be a clear goal that all government entities are aiming towards. Otherwise each agency will continue to do whatever they think makes sense just for them with no convergence in approach and no motivation to work together.
2. Strong Recommendation: Develop and publish work program and roadmap with meaningful measures of progress and success regularly reported publicly on a public data agenda dashboard. NSW Government already have a [public roadmap and dashboard](#) to report progress on their open data agenda.

Whole of government data infrastructure

3. Strong Recommendation: Grow the data.gov.au technical team to at least 5 people to grow the whole of government catalogue and cloud based data hosting infrastructure, to grow functionality in response to data publisher and data user requirements, to provide free technical support and training to agencies, and to regularly engage with data users to grow public confidence in government data. The data.gov.au experience demonstrated that even just a small motivated technical team could greatly assist agencies to start on their data publishing journey to move beyond policy hypothesising into practical implementation. This is not something that can be efficiently or effectively outsourced in my experience.
   • *I note that in the latest report from PM&C, Data61 have been engaged to improve the infrastructure (which looks quite interesting) however, there still needs to be an internal technical capability to work collaboratively with Data61, to support agencies, to ensure what is delivered by contractors meets the technical needs of government, to understand and continually improve the technical needs and landscape of the APS, to contribute meaningfully to programs and initiatives by other agencies, and to ensure the policies and programs of the Public Data Branch are informed by technical realities.*
4. Recommendation: Establish/extend a data infrastructure governance/oversight group with representatives from all major data infrastructure provider agencies including the central public data team to improve alignment of agendas and approaches for a more holistic whole of government approach to all major data infrastructure projects. The group would assess new data functional requirements identified over time, would identify how to best collectively meet the changing data needs of the public sector and would ensure that major data projects apply appropriate principles and policies to enable a data driven public service. This work would also need to be aligned with the work of the Data Champions Network.
5. Recommendation: Map out, publish and keep up to date the data infrastructure landscape to assist agencies in finding and using common platforms.
6. Recommendation: Identify on an ongoing basis publisher needs and provide whole of government solutions where required to support data sharing and publishing (eg - data.gov.au, ABS infrastructure, NationalMap, analytics tools, github and code for automation, whole of gov arrangements for common tools where they provide cost benefits).
7. Recommendation: Create a requirement for New Policy Proposals that any major

data initiatives (particularly analytics projects) also make the data available via accessible APIs to support other uses or publishing of the data.

8. Recommendation: Establish (or build upon existing efforts) an experimental data playground or series of playgrounds for agencies to freely experiment with data, develop skills, trial new tools and approaches to data management, sharing, publishing, analysis and reuse. There are already some sandbox environments available and these could be mapped and updated over time for agencies to easily find and engage with such initiatives.

Grow consumer confidence

9. Strong Recommendation: Build automated data quality indicators into data.gov.au. Public quality indicators provide an easy way to identify quality data, thus reducing the time and effort required by data users to find something useful. This could also support a quality search interface, for instance data users could limit searches to "high quality government data" or choose granular options such as "select data updated this year". See my earlier blog (from PM&C) [draft of basic technical quality indicators](#) which would be implemented quickly, giving data users a basic indication of how usable and useful data is in a consistent automated way. Additional quality indicators including domain specific quality indicators could be implemented in a second or subsequent iteration of the framework.

10. Strong Recommendation: Establish regular public communications and engagement to improve relations with data users, improve perception of agenda and progress and identify areas of data provision to prioritise. Monthly blogging of progress, public access to the agenda roadmap and reporting on progress would all be useful. Silence is generally assumed to mean stagnation, so it is imperative for this agenda to have a strong public profile, which in part relies upon people increasingly using government data.

11. Strong Recommendation: Establish a reasonable funding pool for agencies to apply for when establishing new data infrastructure, when trying to make existing legacy systems more data friendly, and for responding to public data requests in a timely fashion. Agencies should also be able to apply for specialist resource sharing from the central and other agencies for such projects. This will create the capacity to respond to public needs faster and develop skills across the APS.

12. Strong Recommendation: The Australian Government undertake an intensive study to understand the concerns Australians hold relating to the use of their data and develop a new social pact with the public regarding the use and limitations of data.

13. Recommendation: establish a 1-2 year project to support Finance in implementing the data driven recommendations from the Murray Inquiry with 2-3 dedicated technical resources working with relevant agency teams. This will result in regulatory streamlining, improved reporting and analysis across the APS, reduced cost and effort in the regular reporting requirements of government entities and greater reuse of the data generated by government reporting.

14. Recommendation: Establish short program to focus on publishing and reporting progress on some useful high value datasets, applying the Public Data Policy Statement requirements for data publishing. The list of high value datasets could be drawn from the Data Barometer, the Murray Inquiry, existing requests from data.gov.au, and work from PM&C. The effort of determining the MOST high value data to publish has potentially got in the way of actual publishing, so it would be better to use existing analysis and prioritise some data sets but more importantly to establish data by default approach across govt for the kinds of serendipitous

use of data for truly innovation outcomes.

15. Recommendation: Citizen driven privacy - give citizens the option to share data for benefits and simplified services, and a way to access data about themselves.

Grow publisher capacity and motivation

16. Strong Recommendation: Document the benefits for agencies to share data and create better guidance for agencies. There has been a lot of work since the reboot of data.gov.au to educate agencies on the value of publishing data. The value of specialised data sharing and analytics projects is often evident driving those kinds of projects, but traditionally there hasn't been a lot of natural motivations for agencies to publish data, which had the unfortunate result of low levels of data publishing. There is a lot of anecdotal evidence that agencies have saved time and money by publishing data publicly, which have in turn driven greater engagement and improvements in data publishing by agencies. If these examples were better documented (now that there are more resources) and if agencies were given more support in developing holistic public data strategies, we would likely see more data published by agencies.

17. Strong Recommendation: Implement an Agency League Table to show agency performance on publishing or otherwise making government data publicly available. I believe such a league table needs to be carefully designed to include measures that will drive better behaviours in this space. I have previously mapped out a draft league table which ranks agency performance by quantity (number of data resources, weighted by type), quality (see previous note on quality metrics), efficiency (the time and/or money saved in publishing data) and value (a weighted measure of usage and reuse case studies) and would be happy to work with others in re-designing the best approach if useful.

18. Recommendation: Establish regular internal hackfests with tools for agencies to experiment with new approaches to data collection, sharing, publishing and analysis - build on ATO lab, cloud tools, ATO research week, etc.

19. Recommendation: Require data reporting component for New Policy Proposals and new tech projects wherein meaningful data and metrics are identified that will provide intelligence on the progress of the initiative throughout the entire process, not just at the end of the project.

20. Recommendation: Add data principles and API driven and automated data provision to the digital service standard and APSC training.

21. Recommendation: Require public APIs for all government data, appropriately aggregated where required, leveraging common infrastructure where possible.

22. Recommendation: Establish a "policy difference engine" - a policy dashboard that tracks the top 10 or 20 policy objectives for the government of the day which includes meaningful metrics for each policy objective over time. This will enable the discovery of trends, the identification of whether policies are meeting their objectives, and supports an evidence based iterative approach to the policies because the difference made by any tweaks to the policy agenda will be evident.

23. Recommendation: all publicly funded research data to be published publicly, and discoverable on central research data hub with free hosting available for research institutions. There has been a lot of work by ANDS and various research institutions to improve discovery of research data, but a large proportion is still only available behind a paywall or with an education logon. A central repository would reduce the barrier for research organisations to publicly publish their data.

24. Recommendation: Require that major ICT and data initiatives consider cloud

environments for the provision, hosting or analysis of data.

25. Recommendation: Identify and then extend or provide commonly required spatial web services to support agencies in spatially enabling data. Currently individual agencies have to run their own spatial services but it would be much more efficient to have common spatial web services that all agencies could leverage.

Build data drive culture across APS

26. Strong Recommendation: Embed data approaches are considered in all major government investments. For example, if data sensors were built into major infrastructure projects it would create more intelligence about how the infrastructure is used over time. If all major investments included data reporting then perhaps it would be easier to keep projects on time and budget.

27. Recommendation: Establish a whole of government data skills program, not just for specialist skills, but to embed in the entire APS and understanding of data-driven approaches for the public service. This would ideally include mandatory data training for management (in the same way OH&S and procurement are mandatory training). At C is a draft approach that could be taken.

28. Recommendation: Requirement that all government contracts have create new data make that data available to the contracting gov entity under Creative Commons By Attribution only licence so that government funded data is able to published publicly according to government policy. I have seen cases of contracts leaving ownership with companies and then the data not being reusable by government.

29. Recommendation: Real data driven indicators required for all new policies, signed off by data champions group, with data for KPIs publicly available on data.gov.au for public access and to feed policy dashboards. Gov entities must identify existing data to feed KPIs where possible from gov, private sector, community and only propose new data collection where new data is clearly required.

    - *Note: it was good to see a new requirement to include evidence based on data analytics for new policy proposals and to consult with the Data Champions about how data can support new proposals in the recently launched* implementation report on the Public Data Management Report. *However, I believe it needs to go further and require data driven indicators be identified up front and reported against throughout as per the recommendation above. Evidence to support a proposal does not necessarily provide the ongoing evidence to ensure implementation of the proposal is successful or has the intended effect, especially in a rapidly changing environment.*

30. Recommendation: Establish relationships with private sector to identify aggregate data points already used in private sector that could be leveraged by public sector rather. This would be more efficient and accurate then new data collection.

31. Recommendation: Establish or extend a cross agency senior management data champions group with specific responsibilities to oversee the data agenda, sign off on data indicators for NPPs as realistic, provide advice to Government and Finance on data infrastructure proposals across the APS.

32. Recommendation: Investigate the possibilities for improving or building data sharing environments for better sharing data between agencies.

33. Recommendation: Take a distributed and federated approach to linking unit record data. Secure API access to sensitive data would avoid creating a honey pot.

34. Recommendation: Establish data awards as part of annual ICT Awards to include:

most innovative analytics, most useful data infrastructure, best data publisher, best data driven policy.

35. Recommendation: Extend the whole of [government service analytics capability started at the DTO](#) and provide access to all agencies to tap into a whole of government view of how users interact with government services and websites. This function and intelligence, if developed as per the original vision, would provide critical evidence of user needs as well as the impact of changes and useful automated service performance metrics.

36. Recommendation: Support data driven publishing including an XML platform for annual reports and budgets, a requirement for data underpinning all graphs and datavis in gov publications to be published on data.gov.au.

37. Recommendation: develop a whole of government approach to unit record aggregation of sensitive data to get consistency of approach and aggregation.

Implementation recommendations

38. Move the Public Data Branch to an implementation agency - Currently the Public Data Branch sits in the Department of Prime Minister and Cabinet. Considering this Department is a policy entity, the questions arises as to whether it is the right place in the longer term for an agenda which requires a strong implementation capability and focus. Public data infrastructure needs to be run like other whole of government infrastructure and would be better served as part of a broader online services delivery team. Possible options would include one of the shared services hubs, a data specialist agency with a whole of government mandate, or the office of the CTO (Finance) which runs a number of other whole of government services.