

Open Data ThinkTank Government + Industry

2015



THE EVOLUTION OF OPEN DATA STRATEGY AND PRACTICE IN NSW

The Government + Industry ThinkTank is part of the continuing conversation about open data within and between government and industry. Not the same old issues being talked about in the same old way, but an opportunity to challenge assumptions, disrupt existing mindsets and look at the wicked issues standing in the way of open data progress.

The event is a genuine “coming together” of industry and government to have a frank and open conversation, and for leaders and practitioners to have direct input to the development of NSW Government strategy and action plans.

In previous years, the ThinkTank has directly informed and influenced the following NSW Government policies and programs:

- Location Intelligence Strategy
- ICT Strategy
- Open Data Implementation Plan
- data.nsw.gov.au
- Apps4NSW
- NSW Globe

Building on the events from the past 3 years, the Open Data ThinkTank drew input from a mix of government, industry and sectoral representatives with a broad range of experiences and views.

The 2015 ThinkTank featured:

- **A scene-setting opening address from the Minister for Innovation and Better Regulation**, and an update on open data for NSW Government from the Deputy Secretary of the Office of Finance and Services
- **Interactive discussions** highlighting opportunities, challenges and case studies on the use of data to create value and deliver better outcomes to business, government and the community
- **Focused workshop-style sessions** featuring new thinking and initiatives from NESTA, the ABS and CSIRO
- **Opportunities to shape** a new Open Data strategy for NSW

During the day, the participants:

- Talked about emerging trends and fresh initiatives
- Unpacked new opportunities for partnership and development
- Explored new ways of thinking about barriers to progress
- Worked together to bust some common myths and misconceptions about releasing and using open data
- Built a better understanding the value of open data by first appreciating what everyone brings to the table.

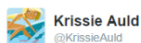
The ThinkTank was opened by Minister for Innovation and Better Regulation, The Hon. Victor Dominello MP.

Drawing on his personal experience using data to make sense of program delivery, key points raised by the Minister included:

- Open data is transforming government: not only the way government and citizens interact, but the way government itself functions.
- There is a strong economic case to fuel data-driven innovation through open data, and by increasing collaboration with the research sector, industry and non-government organisations.
- NSW is home to 64% of Australia's tech start-up businesses, and the NSW government must position itself to secure its share of the value that can be unlocked.
- Positioning open data in the economy, quoting UK Minister for the Cabinet Office, Francis Maude:

'Open data can be a raw material for economic growth – just like iron and coal were to the industrial revolution. It supports the creation of new markets and jobs – businesses of the future, which can help deliver lasting growth.'

The Minister also launched the **NSW Foundation Spatial Data** as web service APIs licensed under Creative Commons Licence By Attribution, the most permissive license available.



Krissie Auld
@KrissieAuld



Follow

@VictorDominello commits to leadership in the open data realm for NSW. Great news!
#ODthink

Minister's opening remarks





New
approaches
and
priorities

William Murphy, Deputy Secretary, Service Innovation and Strategy, Office of Finance & Services provided a perspective on new opportunities, approaches and priorities.

- There is always a need for an innovative approach to open data so that government services and processes can improve and adapt to meet the needs of the community.
- As government, we should embrace open data because it drives innovation in public services and other areas of endeavour, and delivers value, not because a policy says so.
- Industry can help government determine the inherent and potential value of specific government-held data and determine priorities for data release.
- The potential gains in agency efficiencies and customer outcomes more than make the case for allocation of time, skills and resources to open data initiatives.
- NSW Government takes the position of open by default, but it is clear there is still some confusion or barriers to this notion – a key outcome for today is breaking down barriers and myths that slow down the release of open data.
- We need to deal with the fact that our regulations are from another time, and we need to engage executives and business leaders in a conversation about the value proposition for data and prioritising its release.



Allison Hornery
@allisonhornery



Follow

William Murphy from @ICT_NSW calling out the myth that #opendata is a “soft” reform #ODthink

The **Government + Industry panel** was themed **Reality check: open data in NSW for both sides** and featured:

Francisco Urbina

Chair, Spatial Industries Business Association NSW

John Hudson

Director ePlanning, NSW Department of Planning

Dr. Jenny Donovan

General Manager, Strategic Information and Reporting,
NSW Department of Education

Dr Alana Maurushat

Academic co-director of UNSW Cyberspace Law and
Policy Centre + Key Researcher, CRC Data to Decision


The panel shared their perspectives on the current practices and trends in data for public purpose, and identified a series of priority areas of focus for government and industry.

“Data is no longer a by-product:
it **is** the product.”






“
Better data
means you
can start
asking
better
questions
”


Kate Harrington
@KateSHarrington
Data needs to be delivered in a format
that's useful #ODThink


Raul Alberto Caceres
@tototot
"Culture is one of the biggest barriers for
open data sharing" #opendata #ODThink


Cofluence
@Cofluence
Quotable: "If a pic is worth a thousand
words, a map is worth a thousand datasets"
from The Hon @CiscoUrbina #geospatial
#opendata #ODthink


Fred C.
@Mogmog
"US gov released best practice for personal
data masking in open data" by Dr. Alana
Maurushat #ODThink

Key points raised by the panel in their discussion included:

- We need to be drafting policy in a way that anticipates digital delivery as well as data needs and opportunities.
- Government needs to shift from its traditional or gazetted roles – we are rapidly facing the need for a wave of re-thinking about the way we all work: what government does, what industry does and how we work together.
- A focused discussion is needed about re-use because government can't possibly predict how data is going to be used. Engagement with re-user community is critical.
- We need to look more closely at the process of releasing data and then having feedback loops so that new value can be fed back into the cycle.
- The collection of data in Australia is lagging behind, which in turn means the development of new applications, products and services are also lagging behind.
 - For example, the [City of Barcelona](#) has invested in thousands of sensors which creates valuable data intelligence for the city. In addition, this has created a "smart city" platform on which businesses and researchers can build applications. Barcelona has built an economy around how to help other cities be smart cities.
- There are still gaps in understanding what government has by way of data, and the culture of not sharing is still very pervasive. We need to develop better business intelligence tools to analyse data in ways it hasn't been before.
- Having a NSW whole-of-government strategy and policy for open data helps drive this inside agencies – it means it's not just my good idea, everyone thinks it's important.

Des Mooney and **Dr. Steve Woodhouse** from Land and Property Information expanded on the **newly-launched Foundation Spatial Data Web Services** project and released **version 2.0** of the **Foundation Spatial Data Framework** for NSW.



Maurits vd Vlugt
@Mvandervlugt



 Follow

#openData has little use if not discoverable. But NSW LPI data services got 16,000 hits in 1st 24 hrs, with nil advertising! #ODThink



Fred C.
@Mogmog



 Follow

Steve Woodhouse "delivering our open data web services through an elastic infrastructure is critical" #ODThink #Cloud



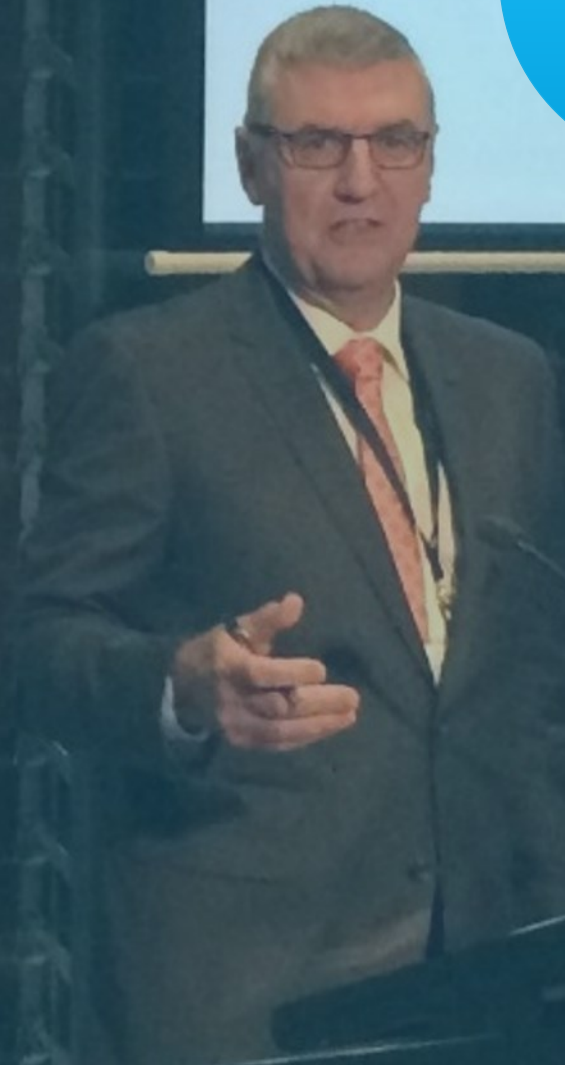
Alex William Cowdery
@cowdery_alex

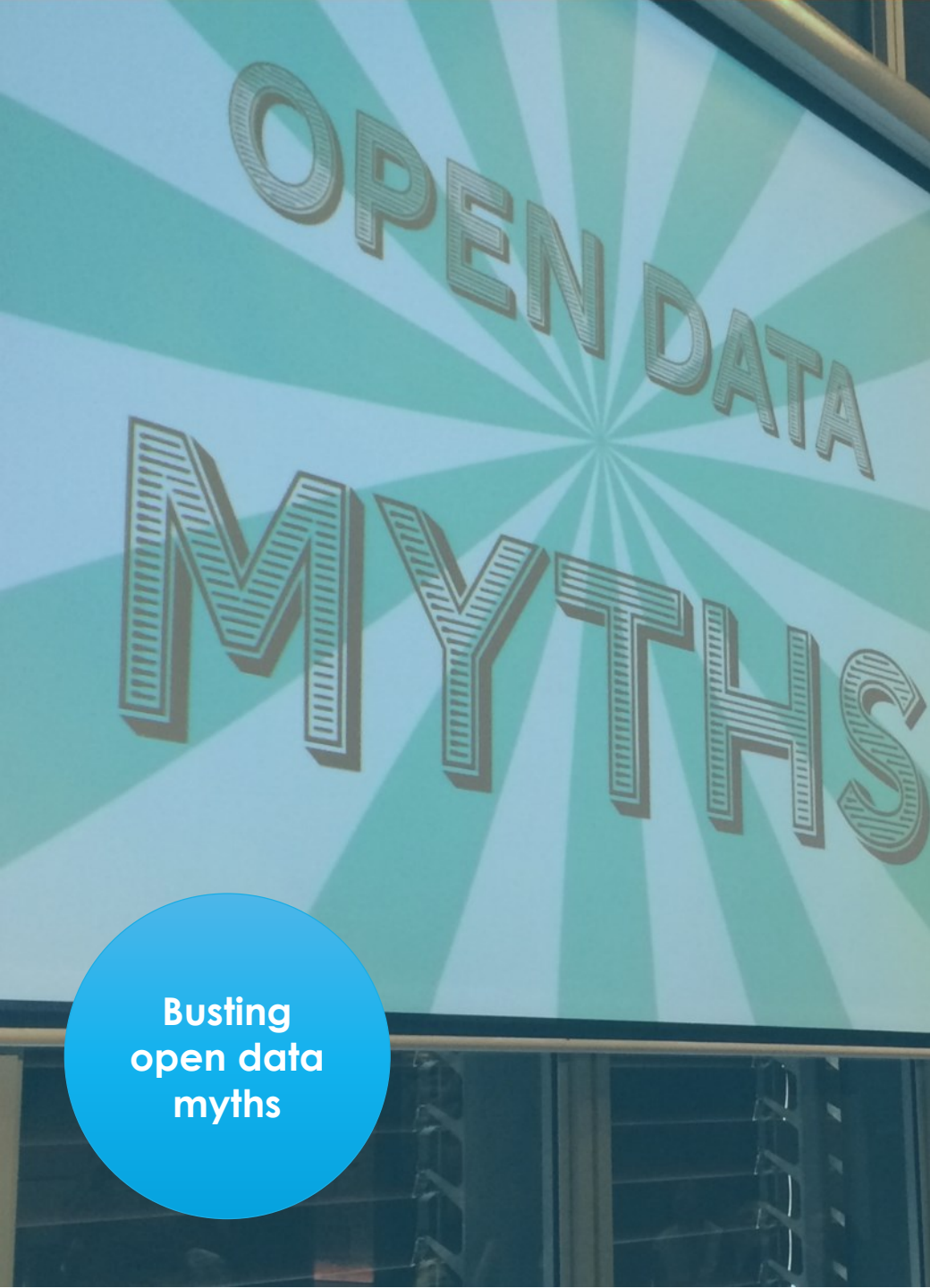


 Follow

#ODThink 9 out of 10 foundation datasets up on the #nswglobe. Well done. #10 coming soon

**Foundation
Spatial
Data**





**Busting
open data
myths**

The **Open Data Myths** session enabled participants to work together to break down open data barriers.

Presented with characters representing nine of the most commonly-held myths around open data, participants were challenged to come up with evidence-based responses to effectively “bust” the myth. **The myths presented were:**

1. **DANGER:** If we release data then people might do dangerous things, such as use the facilities in ways that we can't protect or predict.
2. **CREDIBILITY:** We take our scientific integrity seriously, and this integrity is a major part of the credibility of our agency.
3. **AUTHORITY:** I can't make the decision that I'm the one who is authorised to release data or information.
4. **COST/BENEFIT:** We don't know how the data is going to be used, or even if it will be used.
5. **CONTROL:** People think that if government data is involved, then any app or service that uses the data is a government responsibility.
6. **NEUTRALITY:** If we release any of this data, the government might be perceived to be showing a bias.
7. **THE FRONT PAGE:** If we can't be 100% confident about our data, and if we can't control the message with its release, our leaders may be publicly embarrassed.
8. **INTERPRETATION:** Only people with the necessary skills and subject-matter experience can interpret and then communicate complex data.
9. **USEFULNESS:** We don't have any data that anyone would find interesting or useful.



Each group had **one minute to pitch** their “busting” case to a panel of judges, featuring:

Krissie Auld

NSW Office of Environment and Heritage

Cassie Findlay

Recordkeeping Innovation

Rustum Abdurahman

NSW Centre for Education Statistics & Evaluation



OPEN DATA MYTH #1

DANGER

Geoff

AGENCY WITH
RECREATION DATA



PROBLEM: A lot of people ask us to release data about recreation facilities, however the datasets haven't been checked for 2 years or more. What happens, for example, if campsite information is out of date... such as if the facilities there are burnt, or washed away or not functional? What if people are relying on this information and they are injured or die because the data wasn't accurate?



BELIEF: If we release data then people might do dangerous things, such as use the facilities in ways that we can't protect or predict. We don't want to enable these dangerous acts. So it's better if the data isn't released in the first place. I'd rather not give out the information than worry that people might rely on it.

BUSTED



- Use disclaimers to cover the risk – see Google Maps for example
- Non-government websites have similar data – for example, australiancampsites.com.au
- Other government agencies at state and federal levels have similar data
- The risks of not sharing outweigh the risks of sharing – for example, think hunting season dates
- Use metadata to show how recently the data has been updated
- Communicate a shared responsibility of data with the user – “user beware”
- Users could make the same assumptions based on their previous experience, not your data – for example “there was water there last year”
- Mitigate risks using proven risk management processes and tools
- Courts are less likely to side with stupid people
- Releasing data may actually encourage people to start using facilities they didn't know existed
- Look at new partnerships – for example, Oregon Gov partnering with Trailhead Labs around [data standards for trails](#)
- People will find and trust alternative sources that might be wrong, in the absence of government data
- The private sector is already creating services and tools from this sort of data and they have a much lower risk tolerance than government
- By releasing, you can crowdsource data corrections and additions to improve it

OPEN DATA MYTH #2

CREDIBILITY

Maria
AGENCY WITH
WILDLIFE DATA



PROBLEM: It's been suggested that we release data so that we can crowdsource updates. But this degrades the quality of our data and we don't want to release data that doesn't meet our standards of accuracy and efficacy.



BELIEF: We take our scientific integrity seriously, and this integrity is a major part of the credibility of our agency. We have a system for how we process requested updates to the data, and it works. It may be really slow, but we would rather be slow than release low quality data.

BUSTED



- People understand the difference between 'official' and crowdsourced data and will assess it accordingly providing it is clearly labelled
- Use attribution and share the credibility
- Increase the levels of data available for analysis – by machines, other people, other agencies
- Use a disclaimer in good, plain language
- Credibility and expertise do not reside exclusively with government. You never know what benefits might come from engaging with non government experts, including complementary skills to yours
- Place a value on peer review
- Data can include metadata that helps verify authenticity – for example, GPS coordinates or digital signatures
- Using game theory can help motivate people to contribute to solve problems
- Was accurate \neq is accurate: fast, high-volume data can be cross-checked to help determine validity and authenticity
- Data from the "Internet of Things" can be more reliable than human observation – the larger sample size can help you detect anomalies

OPEN DATA MYTH #3

AUTHORITY

MORE WORK
NEEDED



Carol
AGENCY WITH
TRANSPORT DATA



PROBLEM: The GIPA Act specifies that all government agencies must release information unless there's a strong public interest reason not to. But no person or position is specified as having the authority to release the data. So how do I know if I can or should?



BELIEF: I can't make the decision that I'm the one who is authorised to release data or information. There are no specific guidelines in my agency and I don't want to put myself in the firing line for doing the wrong thing. So I'm more comfortable making a case that it's not in the public interest to release the data.

- It's your job to release the data because open data is NSW Government policy
- Leadership must educate and empower their staff and help them understand that it's OK to release data
- Authority is vested in GIPA
- Create an open data champion inside your agency (and other agencies)

BUT

- There is still a lack of ownership of open data as organisational programs within agencies, so people do not know who to get approvals from
- Clear guidance for staff is also often not there

OPEN DATA MYTH #4

COST / BENEFIT

Andrew
AGENCY WITH
SPATIAL DATA



PROBLEM: There is a lot of data but it's not always easy to release it.

Some of it is in proprietary systems which means we need software and specialist skills to make it suitable for open use. How do we know whether the time, cost and effort will be worth it?



BELIEF: We don't know how the data is going to be used, or even if it will be used. I have to try and make a business case to my executive to invest in creating and then maintaining the data in open, accessible formats without knowing what the usage and benefits will be. I simply don't have the evidence to make a strong case.

**MORE WORK
NEEDED**



- Think differently about costs and benefits
 - COSTS
 - What data is not proprietary? Focus on this as an easy win, and then focus your energies on the other data
 - Document the preparation and release of data as a specification to ensure repeatability
 - BENEFITS
 - Increased use of data and facilities
 - Improved productivity and service delivery
 - Lower costs of feedback
 - Fitness for internal use vs external use
- Examples of new cost/benefit models
 - Emergency services: NSW Emergency Coordination Unit acts as a data brokerage for physical location of hydrants on behalf of councils
 - Enabling decisions by the public – Fires near me, open transport data

BUT

- There are lots of economic arguments (societal benefit, stimulating the economy) but very few financial ones at the agency level so it's still hard to make the case to a CFO
- We need cost/benefit analysis work on open data projects. More of these (perhaps shared on a wiki like data.gov has) would be very useful.

OPEN DATA MYTH #5

CONTROL

Alex
AGENCY WITH
WATER DATA



PROBLEM: We don't know how the data is going to be used. Anyone could take our data and develop an app, or a service or even create an aggregated data source or platform. We see this as being very high risk around security and competition issues.



BELIEF: People think that if government data is involved, then any app or service that uses the data is a government responsibility. If I release open data, I don't know how it will be used and interpreted. If the information is useful to the public, it's better for us to publish it as a PDF so we can retain control over it.

**MORE WORK
NEEDED**



- Put in place appropriate guidelines and policies
- Be transparent about your data
- Leverage the opportunity to use information in real time
- Releasing data optimises its value
- Released data is more cost-effective and gives you a new kind of return on investment
- Releasing data creates and improves collaborative relationships with users and re-users
- Sharing data can help identify and address risks

OPEN DATA MYTH #6

NEUTRALITY

Jean
AGENCY WITH
BUSINESS DATA



PROBLEM: We hold different kinds of data about different kinds of business groups and licensing services. For example, we have data about licensing for builders, but not for accountants or many other types of businesses. We don't want to be seen to show an imbalance in the level of scrutiny of particular groups.



BELIEF: If we release any of this data, the government might be perceived to be showing a bias. If we can't show equal levels of detail for all kinds of transactions or services, then we shouldn't be showing any of them. This allows us to maintain a neutral position.

BUSTED



- If you are transparent, you cannot be accused of bias
- Include a disclaimer with the data
- Implement a rating system, for example like TripAdvisor and PatientOpinion
- Classify complaints so that it's easier to understand the nature of a complaint
- Data can always be improved by releasing it
- User-driven data adds new value
- Assume it's neutral unless proven otherwise
- Aggregate the data
- Releasing data motivates businesses to improve their performance
- Release to allow the crowd to correct misinformation and bias
- Encourage other agencies to consider similar approaches

OPEN DATA MYTH #7

THE FRONT PAGE

Ann
AGENCY WITH
EDUCATION DATA



PROBLEM: We are very anxious about a negative story based on our data appearing on the front page of the newspapers. We aren't confident that our leaders have a high level of tolerance for this kind of risk and we can't be certain that our data will be used and interpreted correctly by the community.



BELIEF: If we can't be 100% confident about our data, and if we can't control the message with its release, our leaders may be publicly embarrassed. It is safer for us to control the narrative based on our own work rather than release the data.

BUSTED



- This argument is based in fear not fact – publishing data may become **good** news
- There is a greater risk in being perceived as secretive
- Releasing data proactively takes the wind out of the sails of would-be gotcha journalists
- GIPA specifics that risk of embarrassment is **not** a reason not to publish data
- By taking charge of the release you can provide more context around it
- Put the purpose of the agency first – in this example it's about education, as well as training, skills and knowledge
- Exposing a need can generate opportunities – particularly enabling the private sector to respond. This has a positive impact on the economy, and reduces the tax burden
- Take the friction out of proving yourself
- Publishing data is part of an improvement cycle - plan to release data from the beginning of your process
- Not publishing data is out of sync with contemporary constituents (think Gen Y) – there is now an expectation that the data is available
- Take the opportunity to share data between agencies to build understanding
- You can't ever be 100% confident in your data – publishing data shows accountability, and welcomes openness

OPEN DATA MYTH #8

INTERPRETATION

Gina
AGENCY WITH
HEALTH DATA



PROBLEM: We understand what our data is telling us, but the public doesn't have the same knowledge and understanding. For example, if we release de-personalised data about cancer clusters that can be geo-located, when matched with data about transport and health facilities the public may panic and interpret certain geographic areas as high risk, and this will have a domino effect.



BELIEF: Only people with the necessary skills and subject-matter experience can interpret and then communicate complex data. This kind of data is safer within the walls of government with the necessary experts, and not in the open domain.

BUSTED



- The internet provides evidence of self-correction by the public of incorrect interpretation of data, for example:
 - Wikipedia
 - Social networking sites like Facebook
 - Google
- More data enables cross-validation, which provides context. For example, Landsat and NASA data overlays help identify flooding areas, which in turn saves lives
- Subject matter experts are everywhere, not just inside your agency – think the research community, specialist businesses as well as other parts of government
- We have the ability to understand who is consuming and wanting to interpret data through the FOI requests
- Data classification helps to strengthen the case that the benefits outweigh the risks
- You can properly identify datasets using metadata, including provenance metadata

OPEN DATA MYTH #9

USEFULNESS

Jack
AGENCY WITH
UTILITY DATA



PROBLEM: We don't understand the demand for any kind of data to be open for access and use. Our agency has a large amount of real-time information about utility usage but this is collected, maintained and used for our own bespoke purposes and we don't see the additional value it may offer to anyone else.



BELIEF: We don't have any data that anyone would find interesting or useful. We collect the data for our own purposes, so there's no need to release it for other uses and find ourselves having to manage a demand that we aren't ready to meet.

BUSTED



- Think Donald Rumsfeld: what are your unknown unknowns?
- Be a sharer: tell people what you have, even if you don't know why they might be interested
- Data could be more insightful to other organisations or people when linked to other datasets – eg utility usage + hardship programs
- What is the cost benefit of not using the data for other purposes? Not releasing the data may actually prevent deploying resources where they are needed – for example, voluntary support around a power outage
- Transparency of the individual pieces of a picture helps to determine the size of a larger problem
- Release as cascading data – this is lower risk to build a business case
- You don't know what value your data may have until you put it out there, some of the results of which could enhance your services



Feature sessions

Two feature sessions on data in practice were presented.

Dr. Kerry Taylor from **CSIRO** presented an overview of the opportunity presented by **Linked Open Data** for public policy.



Cassie Findlay
@CassPF



Follow

Learning about experimental environmental linked open data published by the Bureau of Meteorology lab.environment.data.gov.au #odthink

Matt Berger from the **Australian Bureau of Statistics** presented work on the **Australian Census Longitudinal Dataset** (ACLD) as an example of confidentialisation preserving utility.



Raul Alberto Caceres
@totocol



Follow

So much power that could be exploded from census #opendata #ODThink

Giulio Quaggiotto, Senior Program Manager for Innovation Skills from **NESTA** joined the session from London to talk about big data for public policy.

Giulio shared a number of case studies across a range of thematic areas to demonstrate the innovative ways in which data is being captured, analysed and used by government, industry and citizens.

Giulio's talk included discussion of:

- Are we seeing the end of national statistics as we currently know them?
- Real time feedback loops and the end of surveys
- Government readiness to anticipate short-term trends that appear through big data analysis
- The era of “experimental government”
- Smart cities vs smart citizens

Giulio's closing comment was to suggest that the mature conversation is when we just refer to “data” and avoid the resulting confusion, conflicting focus and fragmentation of parallel discussions about open data and big data.

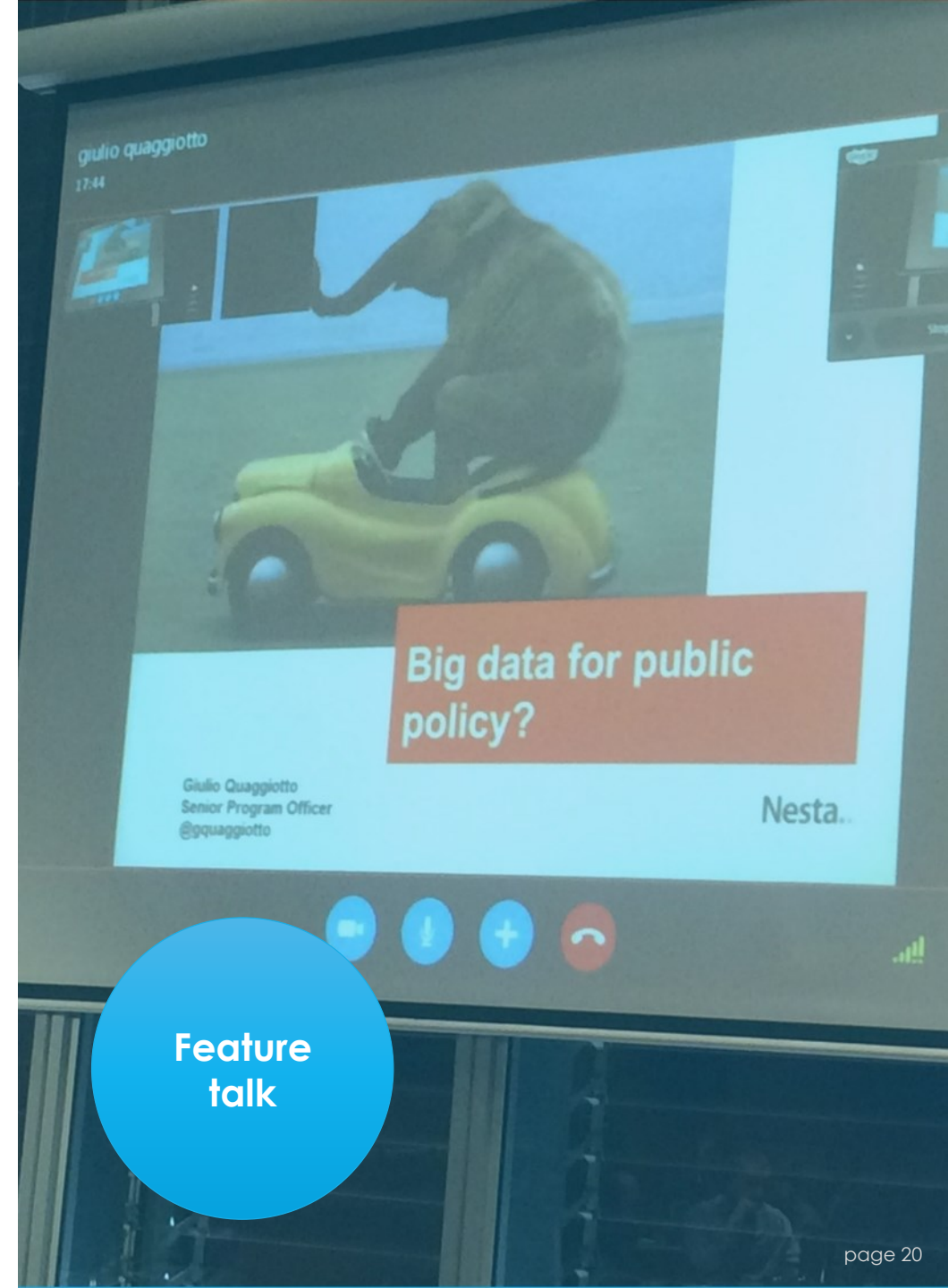


Cassie Findlay
@CassPF



Follow

Interesting comparison of Gallup polls with answers to questions on OKCupid. The latter volumes about 100× bigger datasets
[#ODThink](#)



Feature
talk

The afternoon panel session was themed **Big data in NSW** and featured:

Prof. Mary O'Kane

NSW Chief Scientist and Engineer

Jeremy Moon

Director, Analytics

NSW Procurement, Office of Finance and Services

Mary O'Kane shared her experiences with big data from both an industry and government perspective, looking at issues as diverse as coal seam gas through to aid and development data.

Jeremy Moon presented a demonstration of the new NSW Government procurement visualisation tool. Key issues arising from the demo included:

- Governments being able to see each other's procurement activities starts a new conversation between government and suppliers
- Information presented this way makes it easier to ask the right questions
- Without information like this being openly available across government, let alone outside government, it's difficult to see how well legislation and policy in this area is working





The room debated the issue of whether NSW Government needs an open data strategy, a big data strategy or just a data strategy.

The consensus seemed to point to a data strategy, but with the following points raised as emphasis:

- We should be focusing on what questions we are trying to answer, and let the data follow.
- We need to be asking where are the outcomes and value.
- Government's role in this space is to facilitate, enable and protect where needed. Transparency is key to people making informed decisions.
- We need greater focus on building our capability for visualisation of data.
- The New Zealand approach to Key Result Areas for agencies is a useful model to ensure the focus stays on meeting customer needs, and helps to cut across siloes.



Allison Hornery
@allisonhornery



Follow

“Australians are naturals for data” says
Chief Scientist Mary O’Kane #ODthink

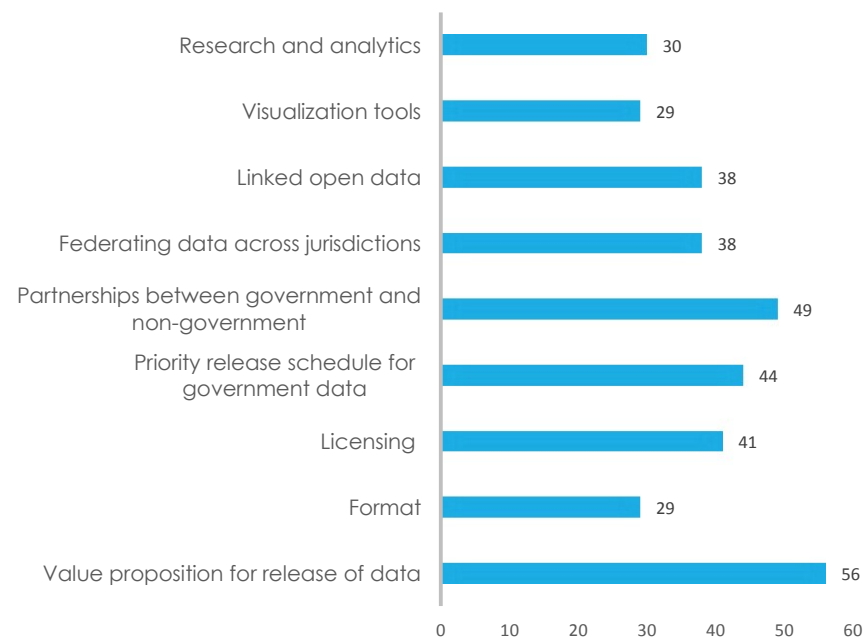
Value and priorities

The final session of the day was led by **Dr. Kate Harrington, Director, Information Strategic Policy at the Office of Finance and Services.**

The focus of this session was to have a conversation about value and priorities in developing open data policies and action plans within government.

As part of the registration process for the ThinkTank, participants were surveyed about what they thought an open data strategy should include.

102 people registered for the ThinkTank, representing a mix of government and industry. The following graph reveals their priorities for a strategy:



We need to
release the right data
that has **high value**
and **high impact**
in the **right way**
as **web services**,
in **open formats**,
with **data quality**
statements
and **openly licensed**
for **re-use**.

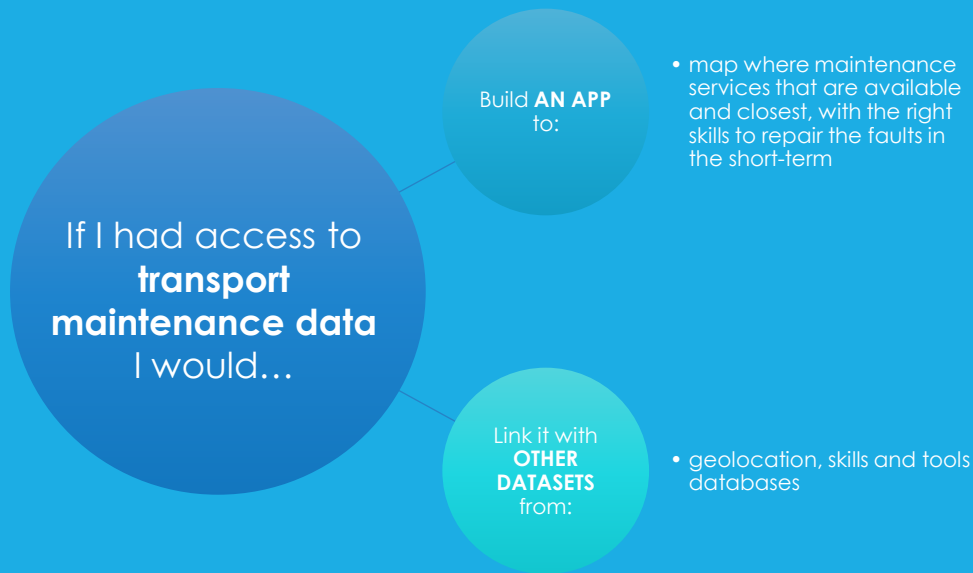
To focus on specific, user-driven outcomes the working session addressed three specific questions:

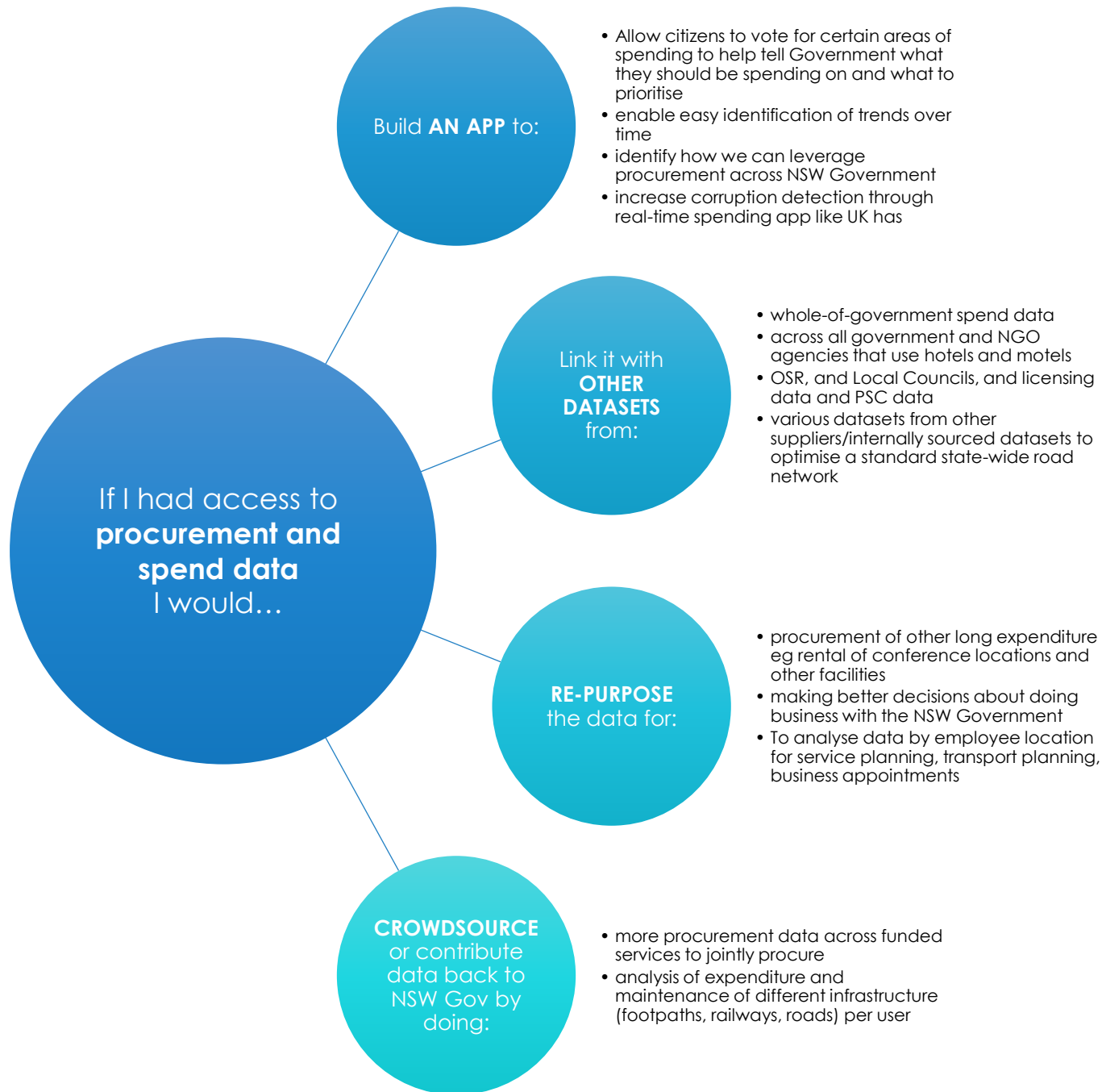
1. **What are examples of “case-makers”** that show the value of releasing certain datasets?
2. **How can both sides work together** to progress the open data program?
3. **What are the specific datasets** that need to be scheduled for priority release, and why?

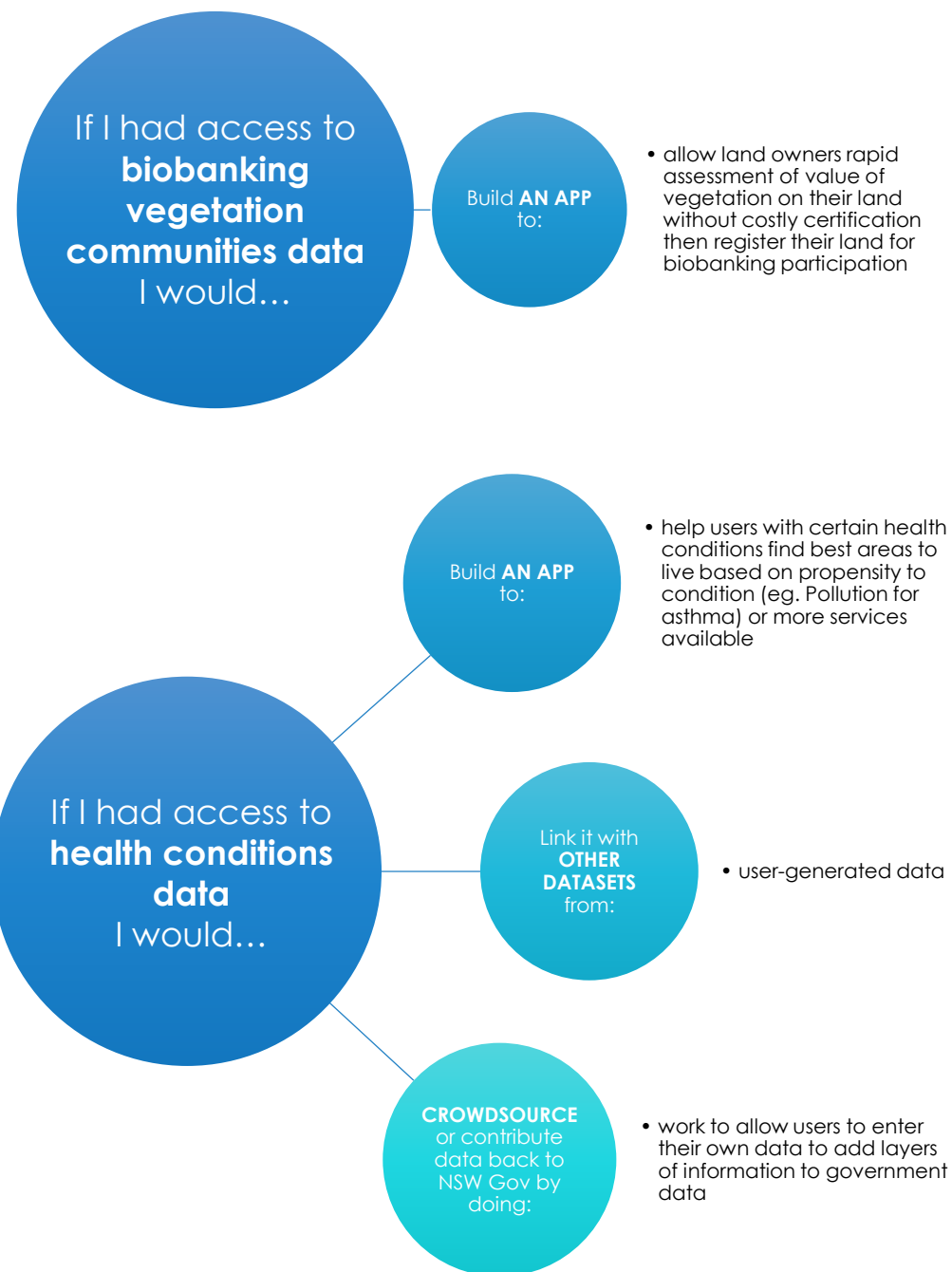
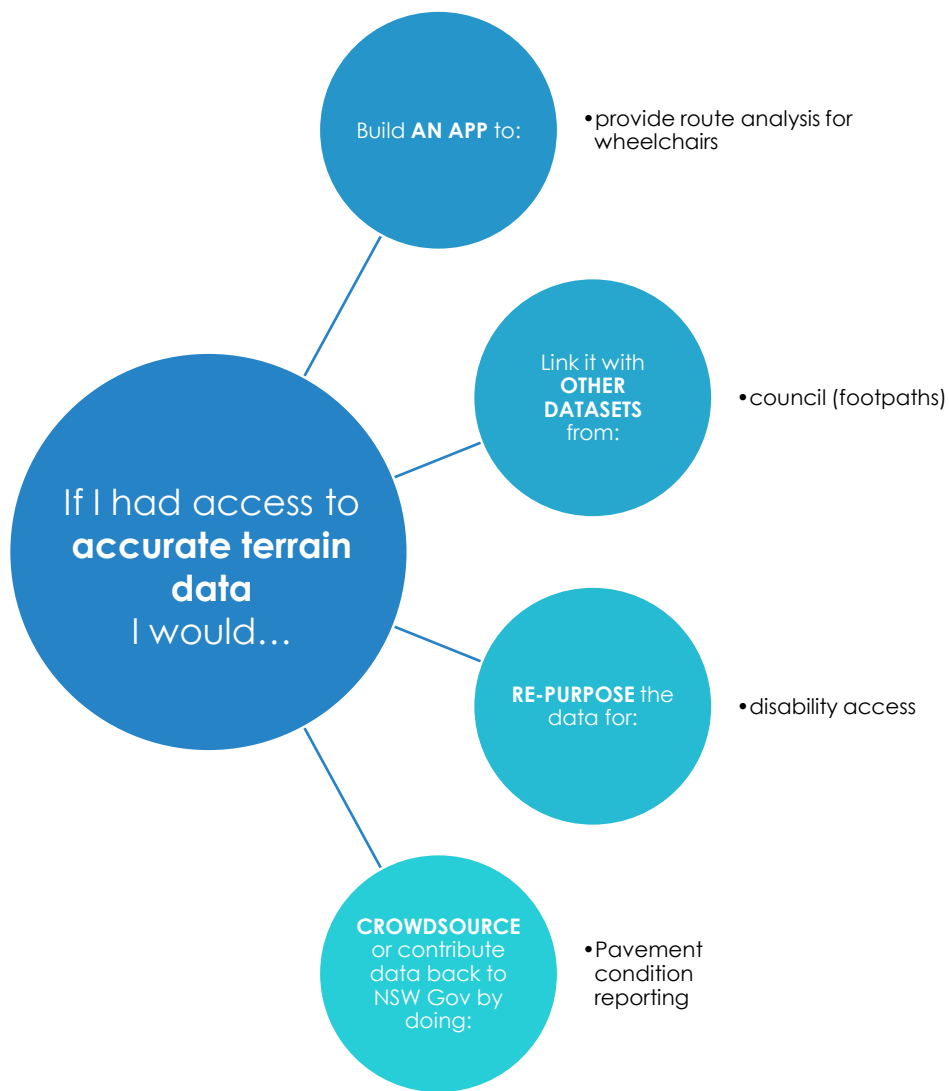
The results from these sessions have been collated and are presented on the following pages.



1. What are examples of “case-makers” that show the value of releasing certain datasets?







If I had access to
**tenancy
complaint data**
I would...

Build **AN APP**
to:

- empower tenants on their rights through outcomes in the Fair Trade Tribunal

If I had access to
**utilities and public
facilities data**
I would...

Build **AN APP**
to:

- enable smart reporting of incidents (vandalism, damaged property) easily instead of the difficult and time-consuming process that currently exists

If I had access to
**business
interactions with
government data**
I would...

RE-PURPOSE
the data for:

- expediting those processes to make it easier and faster to establish and operate a business in NSW

2. How can both sides work together to progress the open data program?

I would like to help with...

- API / Web services specification - how it is presented to developers
- Developing guidance on open data and data.nsw along the lines of data.gov
- Developing practical land management tools
- Giving valuers access to sale and valuations early but return their valuations in the field in real time improving forecasting of property based revenue and better market analysis.
- Government and industry releasing data in an anonymised and confidential way
- Using my association as a vehicle to educate small-medium business/start-ups and connect them into the data.nsw initiatives
- Reviewing the processes in place to facilitate delivery of real-time data stream

I would like to help to:

Cleanse

- Energy
- Water
- Essential services
- Journey
- Property procurement
- Allergy
- Spatial

Link

- All datasets via secure APIs, including as is being done between Service NSW and Police
- Business and industry data
- Business opportunity and indigenous opportunity
- Commonwealth open data, public open data such as the LOD cloud
- Current transport data with ABR, driver times, traffic-congestion modelling
- Disability and mental health data
- Education datasets with transport, affordable/community housing. Create student hubs.
- FACS, health, location, police
- Food composition data to private sector food industry
- Land titles data (Indigenous and Torres Strait Islander title, old system, native title) with land values, ABS, POI services
- Procurement data to NGO leasing data
- Transport data to mobile phone data to improve transport services

Visualise

- Disability and mental health data
- Consumer data
- Government spend
- Landscape health data
- Property leasings

Develop a new service using

- Health related or social impact data
- Landscape health data to prioritise investment in land management
- Property leasing data
- Transport data

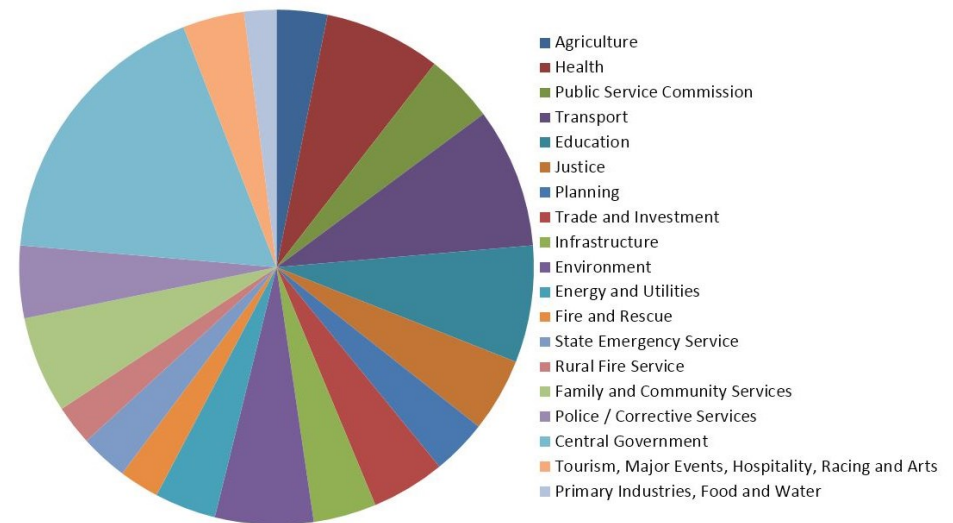
3. What are the specific datasets that need to be scheduled for priority release, and why?

In the lead-up to the ThinkTank, surveys were conducted to identify the key Government clusters and agencies seen as the high priority targets for data release.

The top 5 clusters identified were:

- Central Government
- Transport
- Health
- Education
- Environment

The full set of clusters and agencies identified were:



Taking this set of agencies, participants were asked to identify:

- specific datasets for priority release from these agencies
- the benefits of releasing this data, and
- the challenges associated with releasing the data.

Cluster/agency	Dataset	Challenges in release	Benefits for release
All	Overlaid geo data	Cost; technical barriers	Decision making
Attorney General and Justice	DV incidents	Privacy	Cater resources to meet community needs
Central (Finance & Services)	Procurement spend	Fear of embarrassment to government + executive resistance	Government can make significant savings and clean up waste, non-government organisations can align investment with government, business can make informed decisions about doing business with government
Central	High res. imagery	Cost of infrastructure	Consistent use of imagery and better decision making
Health	Patient time to services and recovery	Privacy	Evaluate health delivery efficiency
Health	Incidence of disease by LGA	Revealing disease clusters	Target health programs by community organisations; stimulate public debate
Health	Patient data	Privacy, security	Patient's complete medical history easily accessed by practitioners
Health	Patient disease data	Privacy	Correlate disease with other demographic/location/agricultural data to determine potential causes
Infrastructure	POIs	Currency	Business feedback updates
Local Government	Parking details	Geo-locating information	Being able to see what parking is available and at what cost
Utilities	Electricity, gas, water	Security, privacy	Reduce service disruption

**HIGH
priority**

Cluster/agency	Dataset	Challenges in release	Benefits for release	HIGH priority
Planning and Environment	Constructions sites/projects	Long-term investigation delays	Regular updates; more frequent release	
Planning and Environment	Road updates	Long discussions, unknown terms and dates	More transparent updates and info; datasets unique formats to introduce	
Planning and Environment	Planning zone information	Accuracy	Better decisions on the opportunity cost of land	
Planning and Environment	Coastal zone boundaries	Currency, controversial data	Better policy and service delivery for community and landowner outcomes	
Planning and Environment	Biobanking sites	Format inconsistency	Participation of landowners	
Police, Juvenile Justice, FACS	Foster children	Privacy, consistency	Holistic understanding of the experience of kids in care	
Primary Industries	Water data	Bureaucracy, CSG denial	Solve water flow issues; ensure water table not effected by CSG	
Transport	Opal	Privacy	Travel routes, demographics	
Transport	Heavy vehicle	Custodian control	Integration and complete dataset for general public and government	
Transport	Speed zones and profiles	Custodian control	Provide accurate and data for 3rd party users and consumers	
Transport	Historic speed and traffic flows	Custodian control	Assist in analysis and understanding of congestion issues on road network	

Cluster/agency	Dataset	Challenges in release	Benefits for release	MEDIUM priority
Planning and Environment	Housing density	Not all data held by planning	Make better decisions for house purchasing	
Planning and Environment	Proposed developments	Only currently provided to LPI and Local Councils	Assist in modelling and predictive analysis for commercial entities	
Planning and Environment	Flood data	Granularity issue, property level of impact on market	Insurance industry; better policy	
Planning and Environment	Areas at risk	Privacy	Identify area at risk and motivate communities to reduce, reuse, recycle	
Planning and Environment	Biosecurity - realtime pest animal and weed data	Cost, resources	Identify emerging risks/threats; more effective eradication	
Transport	Rail/road future building	Culture of fear of backlash	Make better decisions for house purchasing	
Transport	Road sensor data	Revenue, apps from SDPO compromised	Traffic prediction apps	
Transport	Traffic flow (live)	Cost	Real-time traffic data could be incorporated with sat. Nav.	
Utilities	Public energy and utilities (public lighting)	Bureaucratic utilities	Smart lighting - saving energy and money when lights are not required; public safety - ensuring areas like carparks and streets are lit at night	

Cluster/agency	Dataset	Challenges in release	Benefits for release
Attorney General and Justice	Police complaints	Privacy	Identify LAC which feature above average in complaints
Education and Communities	Student test scores by school per student	Embarrassment to schools	Evaluate efficiency of school funding programs
Education and Communities	Student performance by school/region	League table for schools	Change the discussion about targeting education resources; transparency
FACS	Regional population growth	Format inconsistency, many sources of truth, large granularity	Service planning for community development
Health	Local health issues	Privacy	Better understanding of environment factors affecting health of the community
Infrastructure	Social objects	Currency	Improvements of links between government services; self-updates
Police	Safety by LGA	Privacy	Safe communities data across LG areas - trace factors that enable this
Trade and Investment	Business growth	No single source of truth	Attract investment in business and support infrastructure

**MEDIUM
priority**

Cluster/agency	Dataset	Challenges in release	Benefits for release
Health	Hospital presentation in emergency		
Health	Treatment efficacy		
Health	Hospital presentation		
Health	Locations of all health services (including details of what service is provided)	Coordination between agencies	Make it easier for individuals to locate services
Health	Locations of all health services for aged care (including details of what service is provided and vacancies)	Coordination between agencies	Make it easier for individuals to locate services
Planning and Environment	Biobanking sites	Perceived complexity, cost	Build a "ASX" for biobanking info to improve market impact for landowners professionals and property professionals to deliver on legislative goals
Primary Industries	Logistics, production and market data associated with food, agriculture and processing	Privately held data	Evaluate for productivity and efficiency; rural/regional social sustainability
Public Service Commission	GSE success		Understand how successful GSE has been in attracting externals to government
Transport	Historical journey data		
Transport	Maintenance costs of infrastructure		
Transport	Opal	Privacy	Design better public transport routes/schedules/adaptability

**LOW
priority**



Allison Hornery

@allisonhornery

+ Follow

Here's one of the 9 #opendata myths we tried to bust today at #ODthink - great work by busting teams + tough judges!



Social
stream

A number of participants shared their observations from the ThinkTank throughout the day via social media.

These have been curated into a Storify, which can be viewed at <https://storify.com/cofluence/odthink2015>

A photograph of a man in a white shirt and glasses speaking into a microphone at a conference. A blue circle with the word 'Participation' is overlaid on the image.

Participation

Panellists and speakers

The Hon. Victor Dominello MP

Minister for Innovation and Better Regulation

William Murphy, NSW Office of Finance & Services

Francisco Urbina, SIBA NSW

John Hudson, NSW Department of Planning

Dr. Jenny Donovan, NSW Department of Education

Dr. Alana Maurushat, UNSW Cyberspace Law and Policy Centre

Des Mooney, NSW Land and Property Information

Dr. Steve Woodhouse, NSW Land and Property Information

Krissie Auld, NSW Office of Environment and Heritage

Cassie Findlay, Recordkeeping Innovation

Rustum Abdurahman, NSW Centre for Education Statistics & Evaluation

Dr. Kerry Taylor, CSIRO

Matt Berger, ABS

Giulio Quaggiotto, NESTA

Prof. Mary O'Kane, NSW Chief Scientist and Engineer

Jeremy Moon, NSW Office of Finance & Services

Dr. Kate Harrington, NSW Office of Finance & Services



Government representatives

Jeremy Adams

Chris Armstrong

Elise Bates

Oliver Blain

Greg Burgess

Graham Chapman

Ying Chen

Alex Cobb

Emma Collins

Paul Farrell

Siobhan Friis

Simon Gilkes

Lars Hansen

Paula Horeczy

John Hudson

Belinda Jackson

Scott Johnston

Rohini Kannan

Hanno Klahn

Peter Kostantakis

Richard Lehane

Roxane Marcelle-Shaw

David Marcus

Manisha Nanda

Christine O'Keefe

Helen Palmer

Ian Palmer

Lilith Palmer

Wayne Patterson

Daniel Pettman

Rudi Pircher

Karen Purser

Stephanie Salter

Rohita Singh

Penny Webb-Smart

Doug White

Nick Williams



Industry representatives

Kristi Barrow	Susan Hou	Greg Thompson
Anthony Beggs	Andrew Jeffrey	Edmund Tse
Rhys Bittner	Julie Jeltsch	Maurits Van der Vlugt
Melanie Brennan	Paul Kastner	Peter Van Dijk
David Bruce	Roger Kermode	Darius Vitlin
Raul Caceres	Franz Krauth	Matt Wynn Jones
Craig Cashmore	Robert Leigo	Jack Zhao
Judy Chang	Li Liu	
Frederic Chauland	Timothy Mckibbin	
Aimee Chow	Dev Mukherjee	
Marghanita da Cruz	Andrew Perry	
Joe Dantas	Audrey Pulo	
Anne-Marie Elias	Alex Ramsay	
Adam Ezekiel	Nick Randolph	
Eddie Fan	Greg Samuel	
Ed Garvin	Barry Steele	
Daniel Gordon	Amanda Swan	
Dean Hargreaves	Dyana Szibbo	
Peter Hitchiner	Guido Tapia	



OFS Event team

Dr. Kate Harrington

Siobhan Friis

Alex Cobb

Open Data Myth characters

Danger: Peter Kostantakis

Credibility: Rohita Singh

Authority: Paula Horeczy

Cost/Benefit: Matt Berger

Control: Ian Palmer

Neutrality: Siobhan Friis

The Front Page: Helen Palmer

Usefulness: Alex Cobb

Interpretation: Stephanie Salter

ThinkTank co-design and facilitation

Cofluence

Allison Hornery & John Wells



OPEN DATA
MYTH #3
VIRTUALITY

OPEN DATA
MYTH #4
COST / BENEFIT

#ODthink