NSW GOVERNMENT | Finance, Services & Innovation

Evaluation of Data NSW

Report

June 2017

Evaluation of Data NSW

Document number: v1.0	Date: 30 June 2017

Contact details

Name: Kate Cumming	Position: Principal Policy Officer
Business Unit: Information	Division: ICT and Digital Government
Phone: (02) 8276 8188	Mobile: (02) 8276 8188
Email: kate.cumming@finance.nsw.gov.au	

Table of Contents

1. Executive Summary

1.1 Context and Background

1.1.1 Open data

Open data is freely available data that anyone can access, use or share.

The NSW Government is a world-leader in open data. The Government's open data portal, Data NSW, was established in 2009 and was one of the world's first open data portals. Data NSW is still a very active portal, providing access to more than 100,000 datasets and hosting on average 6000-7000 visits per month.

Open data is a strong public good because it is:

- non-rivalrous (use by one party does not reduce its availability to others)
- non-excludable (once available to one party, others cannot be readily excluded from using it)
- economically significant
- transformational, with the potential to generate new careers, drive increased innovation, more efficient government revenues, improved business practices, and better public engagement
- transparent, allowing informed community choices and opening government to scrutiny and opportunities for improvement performance or collaboration. [1]

It is also pre-existing. Releasing open data simply involves the re-use of existing government information assets.

Conservative estimates in Australia place the current direct and indirect value of Australian government open data at up to $25 billion per annum (or 1.5% of GDP), and conservatively place returns on investment in this data over a twenty year period at between AUD 120 and AUD 360 billion. [2]

[1] Bureau of Communications Research (2016). Open government data and why it matters: A critical review of studies on the economic impact of open government data, Department of Communications and the Arts, Australia. Retrieved from https://www.communications.gov.au/departmental-news/open-government-data-and-why-it-matters-now

[2] Gruen, N. Houghton, J. & Tooth, R. (201 4). Open for Business: How Open Data Can Help Achieve the G20 Growth Target, Lateral Economics for Omidyar Network, Australia. Retrieved from www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf, p28

1.1.2 Data NSW

The vision of Data NSW is that the NSW Government aims to make data more accessible to the public and to industry to stimulate innovative approaches to service delivery.

The objective of Data NSW is to make more government data available to the public and increase the flow of data shared internally within government.

The outcomes enabled by the Data NSW program are to:

- improve the transparency and accountability of the NSW Government
- support evidence-based policy development
- enable innovative solutions and service delivery
- improve government services
- empower citizens
- create positive opportunities
- enable government to solve key problems.

To achieve these outcomes, the Data NSW program:

- assists agencies and the public to access, upload, download and request datasets
- facilitates forums, events and meetings to raise awareness of Data NSW
- identifies and addresses inhibitors to open data
- develops blog posts to raise awareness of open data issues, successes and opportunities.

1.2 Evaluation methodology

The broad objectives of the evaluation of Data NSW are to:

- assess the extent to which Data NSW is being implemented as intended
- determine the extent to which the objectives of Data NSW are being met
- determine if Data NSW has contributed to improved outcomes for users
- assess whether Data NSW has delivered value for money.

The evaluation domains and specific evaluation questions are outlined in the table below:

Domains	Evaluation questions
Implementation: Appropriateness	• To what extent does Data NSW address an identified need? • How well does Data NSW align with government and agency priorities?
Implementation: Efficiency	• Is Data NSW being implemented appropriately? • Are users being reached as intended? • How satisfied are users? • Has Data NSW delivered value for money?
Outcomes: Effectiveness	• To what extent has the Data NSW achieved its objectives? • Were there any unintended impacts (positive or negative)?

1.3 Evaluation strengths and limitations

Limitations of the evaluation primarily relate to:

- Limited ongoing administrative data collected from the inception of Data NSW
- Absence of baseline performance data prior to the implementation of Data NSW
- Completeness and accuracy of data
- Timeframes for effective stakeholder consultation
- Small sample sizes in survey exercises.

Strengths of the evaluation are that it was:

- Conducted by business staff actively engaged with the Data NSW portal
- Performed when significant international reviews of government open data portals are underway and so was able to draw from major qualitative and quantitative studies from Australia and overseas
- Able to utliise a wide range of sources to identify a broad range of lessons.

1.4 Key findings: Implementation

The program evaluation shows that does Data NSW does address an identified need and that it does align well with government and agency priorities. The evaluation shows that Data NSW is being implemented appropriately and that it has delivered value for money.

The evaluation shows that more users should be reached and that user satisfaction could be improved.

The wide range of implementation assessments show that Data NSW compares favorably against comparable data portals of its type. It offers comparable functionality with many of the world's leading open data portals. It is well positioned to remain relevant within the wider context of the review of the Open Data Policy and the recently launched NSW Digital Strategy. It can play an important role in responding to stakeholder needs in relation to the Productivity Commission's Inquiry into Data Availability and Use.

The evaluation also shows that key to maintaining the relevancy of Data NSW and improving its implementation is a sustained support from executive and business staff for data quality and data governance as key enablers of digital government. It requires greater awareness of Data NSW, and recognition of the economic potential of platforms like Data NSW that make stable, valuable, standardised and guaranteed sets of government data available for business and community re-use. In practical terms it requires funding strategies and budgets that support the portal's planned and future needs and directions.

Data NSW's role as an aggregator and a means of centralising search for public and private sector users has increased data connectedness and access potential. It has provided a standard approach for any data release, which has saved agencies from having to invent and invest in their own approach. Users like the fact that Data NSW provides a centralised, single point of access to well organised, easily searchable data.

While much has been achieved with portal integration and federation, further seamless integration of Data NSW with other data portals is an area where user experience can be improved. There is also an opportunity to build upon this centralised role, to create larger role as a 'trusted interchange' for data, not just open data, as the evaluation identifies that there is an appetite for safe, secure data sharing.

Key to developing and expanding the remit and implementation of Data NSW will be the alleviation of agency concerns about data risks and championing the inherent value in making data reusable. It will require clear and consistent messaging of the role of Data NSW, how it works and why it is important. Stakeholders must be engaged to build a strong open data culture and multiple channels must be used to share stories about how data can solve problems.

Plain English advice is essential to assist public users and to increase the capacity of public sector open data publishers. Data NSW, like other data portals should include a toolkit with advice on data protection to advise users on how to apply protections relating to privacy and security of customer data.

As in all portals, content is king. As such, it is key learning for implementation that Data NSW must adopt a strategic and coordinated approach to data quality. Data NSW must provide access to data that is primary, timely and easy to use. The amount of government data available on Data NSW must be increased, and it should be fun to play with. For example, users should be able to manipulate the data via visualisations, mappings and a range of other techniques.

The data itself needs to be interesting. The evaluation has shown that publishing high value datasets (such as the NSW Topographic Map) can drive significant interest in and use of government data. Visits to Data NSW are also driven by access to high value data that genuinely enables the development of real solutions to community problems.

The data provided via the portal must be usable. Publishing data in machine readable formats with an open license can greater facilitate its use and re-use. Right now, too many datasets are in PDF form and too few are API-based. NSW must build on recognized standards to foster interoperability, ease of tool development and data re-use.

The evaluation identifies that a significant proportion of NSW government agencies are not sharing data via Data NSW. This stated support for open data initiatives must be transformed into genuine open data practice. This will require a culture change, one driven by collaboration, transparency, supported by strong governance frameworks and embedding open data in system design and contract management. In will mean building open data initiatives as integrated parts of other tools or drivers, such as the Digital Strategy, to minimise impact and maximise benefit for agencies.

This cultural change must understand and respond to users' needs and it means engaging citizens in dataset release prioritisation and design. Comparison with national and international best practice data portals shows that attention is required to improve the Data NSW user design and user experience. This will involve consultation with users to fully understand their needs, and collaboration with other portal owners to share knowledge and lessons learned and will significantly enable further implementation of the open data agenda.

1.5 Key findings: Outcomes

1.5.1 Effectiveness

a) To what extent has the Data NSW achieved its objectives?

The evaluation has identified that Data NSW has partially met its objectives but that further work is required.

Objectives of Data NSW

The objective of Data NSW is to make more government data available to the public and increase the flow of data shared internally within government.

The shorter term expectations of Data NSW was to:

- improve the transparency and accountability of the NSW Government
- support evidence-based policy development

- enable innovative solutions and service delivery
- improve government services
- empower citizens
- create positive opportunities
- enable government to solve key problems.

Metrics and survey data

The current metrics available on Data NSW demonstrate that large volumes of data have been release and that regular downloads are occurring via the platform. This large scale and routine public access of government data is evidence that Data NSW has improved the transparency and accountability of the NSW Government and how citizens access data.

Survey responses however indicated that Data NSW has not yet contributed to the broader objectives of contributing to evidence-based policy and innovative service delivery.

While the survey had a small sample size of six, this does not impact the validity of the findings as the respondents were all key stakeholders of Data NSW and are open data advocates in their organisation.

The following tables list the statements proposed in the survey and respondent answers to them:

Data NSW has supported evidence-based policy development in my organisation		
Yes: 20%	No: 20%	Unsure: 60%

Data NSW has supported innovative solutions and improved service delivery in my organisation		
Yes: 40%	No: 20%	Unsure: 40%

Data NSW has provided positive opportunities in my organisation		
Yes: 20%	No: 20%	Unsure: 60%

Data NSW has empowered agency clients and customers

Yes: 20%	No: 40%	Unsure: 40%

A second survey was distributed to information professionals working in NSW public sector, with 37 responses received. These information professionals are aware of data usage and data strategies in their organisations, but are unlikely to be directly responsible for applying open data strategies in their workplace.

The following tables list the statements proposed in the survey and respondent answers to them:

If Data NSW was not available, would this have an impact on your organisation?		
Yes: 20%	No: 43%	Unsure: 37%

Did the existence of Data NSW help to drive any digital transformation or digital awareness in your organisation?		
Yes: 28%	No: 42%	Unsure: 31%

Areas for improvements

The survey data shows some areas of achievement and areas that can be built on. A 'yes' response was provided for each outcome measurement question which evidence that, to a certain extent, Data NSW's outcomes have been achieved in organisations.

There are clear areas for improvement however in expanding the reach of these outcomes across all agencies.

1.5.2 Were there any unintended impacts (positive or negative)?

a) Results from user surveys and interviews

Survey responses indicated that there are positive impacts of Data NSW, both currently and potentially in the future.

For example, one respondent suggested that Data NSW should play a larger role as a 'trusted interchange' for data. This indicates that Data NSW has a trusted reputation and suggests that there is a growing government appetite for the safe and secure sharing of data.

Another positive impact identified in survey is that Data NSW has provide a standard approach for any data release, which has saved agencies from having to invent and invest in their own approach.

On the other hand, an unintentional impact of Data NSW is that it has contributed to agencies' concerns and anxieties about data sharing. One respondent reported that 'managing the risk of data misuse' was a key challenge with open data, and numerous survey responses pointed to an ongoing concern with open public data. One respondent questioned the rationale behind open data by saying 'Don't understand why Government organisations would want to spend time and energy to hand over data for free to third party developers who then go and make money off it.' This suggests that there are conflicting views around the Open Data Policy principle of "free where appropriate".

Another negative impact relates to confusion between the role of Data NSW and other NSW government data portals. Various survey respondents made comments like 'Get a clear direction/objective of Data NSW' and 'Clearly message why do we need to use Data NSW vs not or other entities'. Other responses counter this however by saying 'My organisation has its own open data portal but Data NSW is more accessible' and 'It is great that all the data is available at a single point of location, easily searchable and well organised.'

Data NSW exists as an aggregator and a means of centralising search for public and private sector users. It does not replace any other portal but it intended to increase connectedness and access potential. The integration back-end of Data NSW should be streamlined to automate as much as possible connection between data portals, and to ensure that data custodians do not have to manually register their datasets on their own portals as well as Data NSW.

b) Areas for improvements

It is clear that Data NSW needs to better articulate and outline its role, purpose and outcomes to both the public and NSW Government agencies. Improving the messaging around the benefits of the portal may facilitate greater usage and participation throughout NSW Government.

It is also clear that there is a need for various technical improvements to make the integration of Data NSW with other data portals as seamless as possible. When Data NSW was first launched, no NSW Government agency had a data portal. The Office of Environment and Heritage, Transport for NSW and Education, among others, have their own portals with various functionalities. While Data NSW does federate and link to these sites, there are opportunities to improve this process.

1.6 Key learnings

The Data NSW program evaluation has provided a wide range of learnings that can be grouped under the following categories:

- User experience
- Quality environment
- Culture and collaboration
- Technical environment
- Budget
- Metrics

This section provides a summary of the key learnings across each of these areas.

1.6.1 User experience

Key lessons relating to user experience are as follows:

- **There are opportunities to improve the interface design**. Data NSW should present a simple and clean interface, with a clearly visible dataset search tool function. The homepage should present less script to provide enhanced user orientation.
- **Data NSW should enable public feedback and contact with data custodians.** Open data frameworks should enable and facilitate an open dialog between users and data custodians.
- **Users should be enabled to better understand data.** Features should be considered that further enable citizens to track and better understand the source of open data, and that enable improved discovery of data insights and related underlying data.

1.6.2 Quality environment

The program evaluation has highlighted the critical importance of the quality environment. Evaluation findings revealed that:

- A large proportion of NSW open data is currently not machine readable. Agencies should be supported to make more of their data available in machine readable formats to allow for its greater reuse potential.
- Data NSW needs data retention and removal policies to ensure third parties can rely on data stability and longevity.

1.6.3 Culture and collaboration

The evaluation identified an ongoing need for cultural change across NSW Government to ensure that the open data objectives are met. The following opportunities were identified:

a) There is ongoing need to build support for open data across government.

A significant proportion of NSW government agencies are not sharing data via Data NSW. Support for open data exists but genuine culture change is still necessary to fully enable open data in NSW government.

Survey results reveal a disconnect between stated open data principles and genuine open data practice. The moderate levels of satisfaction currently reported on in the user surveys could be improved by building greater support and by augmenting the site's technical features.

The Information Commissioner provides current high level support as NSW Open Data Advocate. This support should be leveraged to enable open data and drive organisational open data support at the executive and business levels. This will help to raise awareness of open data and will help to build strong governmental drivers for open data.

It is also important for NSW to operate as part of a national data framework and to build data frameworks that are interoperable with federal, state and local government frameworks. This level of interoperability delivers the greatest value and return on investment to the community.

b) Agencies should be enabled to open data.

The NSW Government should develop a toolkit to consolidate its existing open data guidance material and to provide agencies with clear steps on how to publish open data. This will minimise unnecessary queries to the Data NSW website.

In addition, sharing of knowledge via agency collaboration forums should be encouraged. Key open data stakeholders have significant insights to share on open data enablers and inhibitors that can benefit all open data initiatives.

Data must be used to solve problems and sharing these stories will also help to build a strong open data culture.

c) Assistance should be provided to users to understand and use open data

Advice should be developed to help citizens understand what open data is and how it can be used. Regular surveys or consultation should also occur with the public to identify that their open data needs are and to determine how these needs can be met. The community should also be engaged in dataset release prioritisation and design. Publicly listing all government datasets and committing to ongoing forward release as shown by the data publication schedules on DataSF can also help the community to understand and use open data.

Data releases should be promoted in ways that engage the community and that share data-driven insights with the widest possible audience to build a strong and collaborative open data culture. This should include engagement with social media.

In addition the data prioritised for release should reflect genuine community needs.

1.6.4 Technical environment

a) Data visualisation should be supported in Data NSW

Metadata and data storage layers should be designed to enable data visualisation to be used as much as possible, so that users can easily and quickly get an understanding about the dataset, and decide whether it is worth continuing with a more detailed analysis.

b) Data NSW should enable open data by design

The Data NSW government holds is often locked into inflexible IT systems and retrieving the data can be a costly exercise requiring detailed business case or contractual amendment. Better, flexible and more intelligent IT systems supported by transparent contracts are essential. Future IT contracts must allow for easy and uninterrupted access to data held on the behalf of NSW government agencies.

c) Data NSW should utilise APIs

API-enabled search functionality should be used to augment the current search functionality of Data NSW. API access to high value datasets should also be enabled wherever possible.

1.6.5 Budget

Data NSW has delivered value for money. Data NSW runs on minimal costs but has still released very substantial amounts of data to the public and still drives active data publishing and data download activities every month.

However, funding strategies and budgets should be developed to support the portal's planned and future needs and directions, not just technical and maintenance requirements. The evaluation revealed that Data NSW needs to embed ongoing and adequate budget allocations into its operations, in order to best meet its goals, provide a stable resource for innovation and to adapt to the challenges it faces.

1.6.6 Metrics

Open data is meaningful if it is used. There opportunities to improve efforts to develop meaningful measures and indicators that try to track the use and impact of open data.

1.7 Recommendations

The following key recommendations are made based on the research and learnings represented in this program evaluation. Implementing these recommendations will help Data NSW to further achieve its objectives, increase user satisfaction and deliver value for money.

1. Identify the work required to improve the Data NSW user experience, focussing on interface requirements, streamlined approaches to data publication, improved public contact with data custodians and tools to enable users to better understand data.

2. Identify the necessary components of a data quality framework in NSW, and identify best practice national and international resources that could be adapted or adopted as part of this framework.

3. Work with NSW government agencies to identify the drivers and inhibitors of open data release in NSW.

4. Develop strategies based on identified data release drivers and inhibitors to improve open data release in NSW government.

5. Utilise social media to engage the community and public sector in open data.

6. Undertake research into the open data frameworks, metadata standards and quality tools required to support further data visualisation, spatial enablement and the increased development of APIs.

7. Investigate options for sustainable and ongoing funding options for Data NSW.

2. Introduction

This report presents the results of an evaluation conducted by the NSW Department of Finance, Services and Innovation on Data NSW.

The evaluation was performed by members of the Information team, Policy and Innovation, Information and Digital Government Division DFSI. The review was undertaken over the period of May to June 2017.

This introduction examines:

- Data NSW and open data
- the value of open data
- the functionality and program logic of Data NSW.

2.1 Data NSW and open data

Data NSW is a platform for sharing NSW Government open data.

Open data is freely available data that anyone can access, use or share. To be open data, data must be:

- in the public domain and provided under an open license which enables the data's use and reuse
- easily shared, used and talked about
- available in a standard, structured format, so that it can be easily read and processed by a machine
- guaranteed to be available and consistent over time so that it can be relied on, and so that tools built using it will remain operational
- traceable, so that its origins are known and its contents can be trusted. [3]

Governments generate vast amounts of data in the course of service provision and other business operations. Governments across the world make portions of this data (generally data that contains no personal or sensitive information) available as open data. Government open data initiatives have generally been motivated by transparency and accountability drivers, and concerns to make data available as a public good.

[3] https://theodi.org/guides/what-open-data

Open government data can create many direct and indirect benefits to government, industry and the broader community:

Benefits to government:	• improvements in operational efficiency • savings in overhead expenses through reuse and sharing • improved engagement with stakeholders
Benefits to industry	• new product and service creation • existing service improvement • increased business growth and innovation
Benefits to the community	• enhanced engagement • increased choice and informed decision-making

2.2 Open data portals

Open data portals categorise and provide access to open data. Open data portals provide keyword search and browsing interfaces that enable users to find relevant datasets. Metadata on open data portals describes the organisations making the data available, as well as the geographical, jurisdictional, temporal and technical context of the data.

Open data portals provide mechanisms for users to be able to access and download open data and outline the license conditions that govern data access and use.

Data NSW is NSW government's open data portal for finding, sharing and using NSW Government open data.

2.3 The value of open data

The use and reuse of government open data can help grow the economy, innovate business, improve service delivery, enhance transparency and transform policy outcomes. Recent studies have sought to identify and quantify the specific benefits that open data can bring.

A 2013 McKinsey Global Institute report, *Open data: Unlocking innovation and performance with liquid information* [4], estimated that globally, the annual economic

[4] Manyika, J. Chui, M. Farrell, D. Van Kuiken, S. Groves, P. Doshi, E. (2013) Open Data: Unlocking innovation and performance with liquid information. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/business-functions/digital-

potential of open data could be between USD 3-5 trillion dollars. The potential value would be divided roughly between the United States ($1.1 trillion), Europe ($900 billion) and the rest of the world ($1.7 trillion).

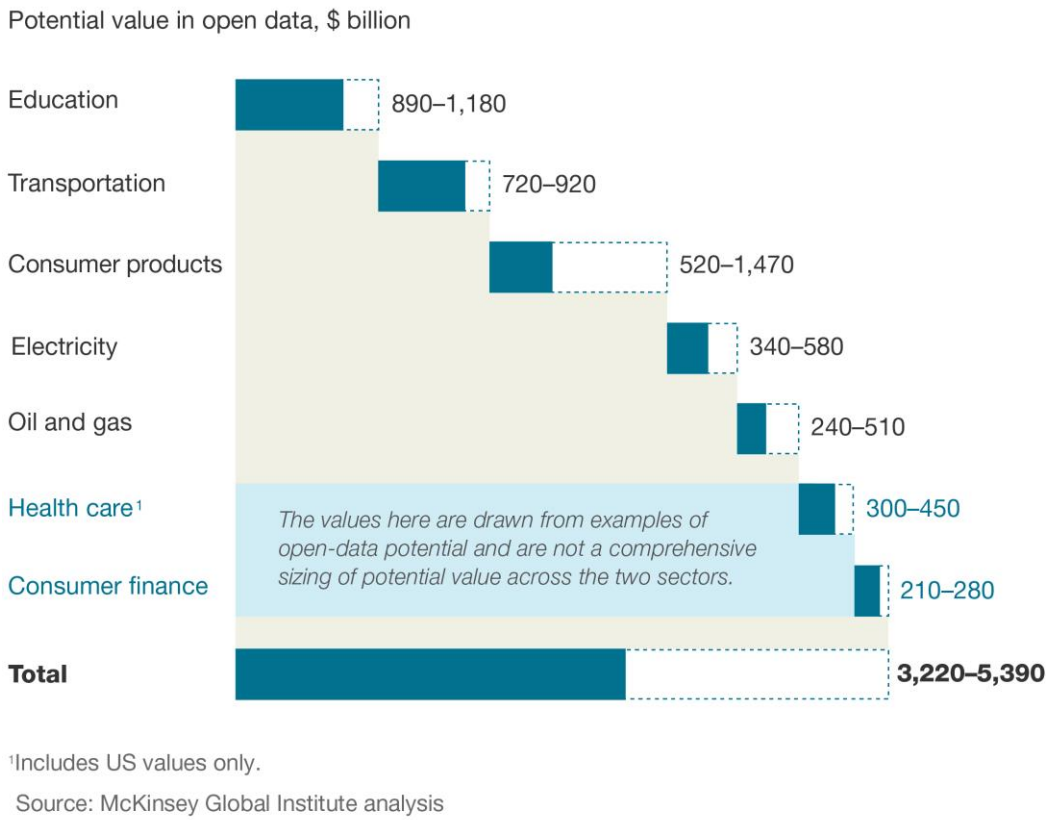


Figure: Annual potential value of open data across seven key domains ($USD3 - $5 trillion)

The McKinsey report says that while open data has already contributed to the development of many new data-driven businesses, and has helped existing businesses to better understand markets, develop new products and services and identify efficiencies, its future potential is very significant:

> Open data can become an instrument for breaking down information gaps across industries, allowing companies to share benchmarks and spread practices that raise productivity. Blended with proprietary datasets, it can propel innovation and help organisations replace traditional and intuitive decision-making processes with data-driven ones.[5]

The McKinsey report identifies the following sources of value realisation through the release of open data.

Open data enables:	**Which drives potential outcomes:**

mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information It should be noted that while the McKinsey report says government open data is the largest and most prominent form of open data, its analysis also includes other forms of open data.

[5] http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

Benchmarking	• exposes variability • promotes transparency • fosters business improvement
Opening of previously unavailable data	• improves the process of defining and creating products and services • develops better products and services
Transparency	• fosters competitiveness • creates opportunities to better match supply and demand • raises the quality of decisions by giving citizens and consumers more tools to scrutinize business and government • enables improved consumer choice • enables cost savings • improves convenience • drives changes in behaviour for consumers, business and government
Data liquidity	• enables new opportunities for large-scale collaboration among individuals, companies, governments, and other organizations [6]

A 2014 Australian report, *Open for Business: How Open Data Can Help Achieve the G20 Growth Target* [7], identifies that open data could contribute half of the G20's economic growth target of 2% (as agreed during the G20 finance ministers and central bank governors meeting in Sydney in February 2014).

The report suggests that McKinsey report's estimates of open data value were possibly conservative, as the methodologies used in economic modelling traditionally underestimate the revolutionary industry changes and order of magnitude benefits that transformations in the digital economy can generate.

[6] Manyika, J. Chui, M. Farrell, D. Van Kuiken, S. Groves, P. Doshi, E. (2013) Open Data: Unlocking innovation and performance with liquid information. Page 7

[7] Gruen, N. Houghton, J. & Tooth, R. (2014). Open for Business: How Open Data Can Help Achieve the G20 Growth Target, Lateral Economics for Omidyar Network, Australia. Retrieved from www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf

The authors estimate that the current direct and indirect value of government data in Australia is up to $25 billion per annum. Focussing specifically on open government data, the report estimates that returns on investment in Australian government and research data over a twenty year period will range between AUD 120 and AUD 360 billion. The authors also state that 'these estimates are probably conservative'. [8]

Their modelling indicates that doubling accessibility and use of open government data would add 0.27% of Australia's cumulative GDP to the economy over the next five years. The authors note that combining government data with other forms of open data in research and business domains will significantly increase this figure.

The report also flags that the release of open data is a unique form of micro-economic reform as it has little to no negative consequences for any group or sector of the community and simply involves the re-use of existing information assets.

The authors state that while some of this projected value is already being obtained, much more can and should be realised through adopting a more aggressive open data agenda. To realise the benefit of open data, the report recommends that governments:

- intensify the release of existing public sector data, both government and publicly funded research data
- mandate the disclosure of information such as corporate financial reporting data
- induce greater data release through influencing information architectures in certain industries such as mining, or in industries where government funds service delivery, such as health or education
- contribute to the develop of standards, including public reporting standards that enable better benchmarking and that create inducements for data release.

In February 2016, the Bureau of Communications Research in the federal Department of Communications and the Arts published *Open government data and why it matters: A critical review of studies on the economic impact of open government data*. [9]

This report provides a strong rationale for governments to take a default position of making government data more accessible. It highlights that open government data is a strong public good as it is:

- non-rivalrous (use by one party does not reduce its availability to others)
- non-excludable (once available to one party, others cannot be readily excluded from using it)

[8] Gruen, N. Houghton, J. & Tooth, R. (201 4). Open for Business: How Open Data Can Help Achieve the G20 Growth Target, Lateral Economics for Omidyar Network, Australia. Retrieved from www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf, p28

[9] Bureau of Communications Research (2016). Open government data and why it matters: A critical review of studies on the economic impact of open government data, Department of Communications and the Arts, Australia. Retrieved from https://www.communications.gov.au/departmental-news/open-government-data-and-why-it-matters-now

- economically significant and able to generate up to $25 billion per year, or 1.5% of Australia's GDP
- transformational, with the potential to generate new careers, drive increased innovation, more efficient government revenues, improved business practices, and better public engagement
- transparent, allowing informed community choices and opening government to scrutiny and opportunities for improvement performance or collaboration.

To provide maximum public benefit the report argues that governments should:

- significantly encourage the use and re-use of open data
- provide open data at no cost, or at most, priced at the short-run marginal cost of making it publicly available provide raw government data in a machine readable format using open standards
- release high value datasets across governments, identified in the report as including spatial data, health data, transport data, mining data, environmental data, demographics data and real-time emergency data. [10]

There is increasing evidence therefore that:

- there is significant potential value in open government data
- benefits to the community, industry and to government can be realised through open data release
- governments should work to encourage the use and re-use of open data
- governments should identify and release high value datasets
- governments should work with industry to enable open data to be released by and shared across business, government and community sectors.

2.4 The evolution of Data NSW

Data NSW was first established in 2009. Internationally, it was the first state based open data portal in the world and it was among the world's first government open data portals.

[10] Bureau of Communications Research (2016). Open government data and why it matters: A critical review of studies on the economic impact of open government data, Department of Communications and the Arts, Australia. Retrieved from https://www.communications.gov.au/departmental-news/open-government-data-and-why-it-matters-now, pp36-37

2.4.1 International open data context, 2009 – 2017

Across the world in 2009, there was significant momentum around the release of open government data. The world's first government open data portal, data.gov, was established in the United States in May 2009. Following this data.gov.uk was established in September 2009 and in December 2009 the Commonwealth government released its *Government 2.0 Taskforce Report.* This report recommended that the federal government 'make public sector information open, accessible and reuseable' and led to the establishment of data.gov.au in 2010. [11]

Government open data initiatives have grown substantially since this time. Each year the World Wide Web Foundation issues a report to track and assess open data portals and initiatives across the world. The 2017 *Open Data Barometer Report* [12] was published in May. It identified that in 2016-17, there were data portals in 115 countries around the world. This represents a 25% growth in countries with open data portals and agendas compared to the 2015 survey. This indicates that open data continues to be a growing international initiative supported by more than half the countries in the world, including all G20 nations.

It is not just government that is investing in open data. Major companies have over the last five years made financial and infrastructure investments to support the use of open government data. Some examples are:

Organisation	Established	Service	Description
Google	2010, expanded 2016	Public Data Explorer	Platform for exploring, visualising and communicating about large public datasets [13]
Google	2013	Funding open government programs	AUD 5 million in grants for programs that support open government data [14]
Microsoft	2014	Open Government Data Initiative DataLab	Cloud-based open data catalogue that provides access to government data and enables developers to obtain data via open standards Application Programming Interfaces (API) [15]

[11] http://www.finance.gov.au/sites/default/files/Government20TaskforceReport.pdf?v=1
[12] Open Data Barometer 4th Edition — Global Report, May 2017 | The World Wide Web Foundation http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf
[13] https://www.google.com/publicdata/directory
[14] https://publicpolicy.googleblog.com/2013/01/promoting-civic-innovation-through.html
[15] http://ogdifrance.cloudapp.net/Home/About

Intel	2014	Data Services Accelerator	Innovation pipeline for supporting solutions that use open government data to develop services [16]
New York University	2014	OpenData500	International study of major companies using government data to generate new business, develop new products and services or to create social value. [17] The Australian version of OpenData500 was founded in 2015. [18]

2.4.2 NSW context, 2009 – 2013

Internationally, the first government hackathon, Apps for Democracy, was held in September 2008 in Washington DC. [19] In 2009 Professor Mary O'Kane, NSW Chief Scientist, partnered with the Department of Finance and Services (DFS) (now DFSI) to develop and launch the apps4NSW program. Data NSW was established in 2009 as a platform to support the apps4NSW program.

Apps4NSW was a series of competitions run by DFS. It challenged developers to use NSW government open data to create innovative web and mobile applications. Each apps4NSW competition was based on a theme and around specific collections of data. In each competition individuals or teams pitched concepts to a judging panel to win a cash prize (ranging between $15,000 - $50,000) and support to develop their proposed concept. The objective of the apps4nsw competitions were to use public sector data to capture innovative ideas useful to the people of NSW. The competitions were a channel for engaging with the public, industry and government employees to improve the delivery of government services using ICT. [20] The apps4NSW competitions ran between 2010 and 2015.

Data NSW was built to support the first apps4NSW challenge in 2010. To support and provide content for this first competition, DFS staff searched and scraped data from state government websites and the first Data NSW platform was built to house this data. Following the first apps4NSW competition, DFS staff conducted annual scrapes of government websites to add extra or updated data to the Data NSW catalogue.

[16] http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/intel-data-services-accelerator-brief.pdf
[17] http://www.opendata500.com/
[18] http://www.opendata500.com/au/
[19] A description of Apps for Democracy: An Innovation Contest is available at https://isl.co/work/apps-for-democracy-contest/
[20] https://data.nsw.gov.au/apps4nsw

In this period the Data NSW website was simply a catalogue list of datasets. In 2011 the site was given its first refresh. Data themes and storytelling were added, and the site was upgraded to be more than the initial simple listing of datasets.

By this stage Apps4NSW was recognised for making a significant contribution to open government, open data and technology innovation. Between 2010 and 2011, significant apps and innovations had been developed through the apps4NSW challenges. *Go Play*, a school holiday activity app, *Demographic Drapes*, an app enabling access to Australian Bureau of Statistics data, and the Sydney Buses SMS service, an early precursor to real-time transport apps, were developed based on open government data. [21] In the 2011 Premier's Awards recognising excellence in public services, apps4NSW won the award for Innovation in Technology [22] and Data NSW was a key component of the apps4NSW infrastructure.

2.4.3 NSW context, 2013 – 2017

Between 2012 and 2013 the popularity of the apps4NSW competitions began to wane and Data NSW began to be recognised as a key government asset in its own right. In 2013, the first NSW Open Data Policy was established. [23] This policy was developed to help agencies to embed open data principles in their operations. The policy identified Data NSW as an open data portal for all NSW government agencies to use, to help improve the way they interact with customers. [24] Following the promulgation of this policy, Data NSW transitioned from being a support environment for apps4NSW to become an environment to make government data more accessible to the public and to industry, and to stimulate innovative approaches to service delivery.

Data NSW was refreshed in 2013 to coincide with the release of the Open Data Policy. Functionally, its refresh aimed to:

- Make dataset access easier
- Federate with Commonwealth, Queensland and South Australian data portals
- Create agency-based context linkages to datasets
- Enable responsiveness on mobile devices
- Establish social media connectivity
- Connect the site more specifically to the Open Data Policy.

Strategically, the key outcomes of the Data NSW 2013 refresh were to improve the transparency and accountability of the NSW Government, support evidence-based policy development, enable innovative solutions and service delivery, improve government

[21] The full list of products developed through the apps4NSW competitions is available at https://data.nsw.gov.au/apps
[22] http://www.publicserviceawards.nsw.gov.au/ArticleDocuments/164/2011%20Premiers%20Public%20Sector%20Awards%20Hall%20of%20Fame.pdf.aspx
[23] https://www.finance.nsw.gov.au/ict/sites/default/files/resources/NSW_Government_Open_Data_Policy_2016.pdf
[24] https://www.nsw.gov.au/news-and-events/news/nsw-government-delivers-open-data-policy/

services, empower citizens, create positive opportunities and enable government to solve key problems. [25]

In 2013 Data NSW also moved to the CKAN platform, an open source data platform and management system widely used internationally to support and federate/connect open data portals. CKAN provides tools for data publishing, sharing, searching and management. It also provides a set of visualisation tools, such as interactive tables, graphs and maps. There is also a native API for querying and retrieving data. [26]

In a media release about the Data NSW re-launch, Minister for Finance and Services Dominic Perrottet said 'the government is committed to making government data more readily available and identified open data as a key driver for the NSW digital economy'. [27] In 2014 open data became a Premier's Innovation Initiative, and non-government sectors were asked to suggest new and creative ideas about what could be done with potential datasets and to fast track the release of open data. [28]

Since 2013 significant work has been undertaken to federate NSW agency-based based data portals and state and federal data portals with Data NSW. The NSW agency-based portals that have been federated with Data NSW are:

- SEED – NSW Environmental Data Portal, providing access to datasets from NSW Resources and Energy, Department of Planning and Environment, NSW Office of Environment and Heritage, NSW Environment Protection Authority, NSW Department of Primary Industries – Water, WaterNSW, NSW Spatial Services [29]
- HealthStats – Ministry of Health [30]
- Cancer Institute NSW [31]
- Bureau of Health Information [32]
- Education – Centre for Education Statistics and Evaluation (CESE) [33]
- Office of Environment and Heritage [34]
- Transport (Performance and Analytics) [35]
- Transport Open Data Portal [36]

[25] https://www.finance.nsw.gov.au/ict/sites/default/files/resources/Open_Data_Policy.pdf
[26] https://ckan.org/
[27] https://www.nsw.gov.au/news-and-events/news/revamped-open-data-portal-open-to-the-public/
[28] https://www.nsw.gov.au/news-and-events/news/premiers-innovation-initiative-expressions-of-interest-open/
[29] https://www.seed.nsw.gov.au/
[30] http://www.healthstats.nsw.gov.au/
[31] http://www.statistics.cancerinstitute.org.au/
[32] http://www.bhi.nsw.gov.au/
[33] http://www.cese.nsw.gov.au/
[34] http://data.environment.nsw.gov.au/
[35] http://www.bts.nsw.gov.au/
[36] https://opendata.transport.nsw.gov.au/

In 2016, the NSW Open Data Policy was refreshed to support new government initiatives relating to the Information and Privacy Commissioner's role as NSW Open Data Advocate and the establishment of the NSW Data Analytics Centre. The refreshed open data policy outlined six open data principles. These state that agency open data must be:

- Open by default, protected when required
- Prioritised, discoverable and useable
- Primary and timely
- Well managed, trusted, authoritative
- Free when appropriate
- Subject to public input. [37]

As at May 2017, 57 government organisations were publishing open data through Data NSW. [38] Ten of these organisations are federal or local government bodies, meaning 47 NSW government agencies are publishing open data through Data NSW. A full list of all organisations publishing data through Data NSW is at Appendix 3.

The evolution of Data NSW therefore shows that it was initially established as a platform for a program, apps4NSW, and served this purpose between 2009 and 2013. In 2013 it was reinvented as a platform for improved service delivery, government and community collaboration, service innovation and open government transparency.

The 2016 NSW Government Open Data Policy says that agencies must link their 'datasets to Data NSW for discoverability and availability'. [39] Support for Data NSW has however principally been driven by a coalition of the willing, primarily those agencies that are engaged with open data and that have advanced data maturity practices. There is a need to engage with agencies that have not yet widely embraced open data initiatives, or that are not significantly advanced with open data, to expand the data that is publicly open and accessible via the Data NSW portal and to start unlocking the potential of government open data for the community, industry and government itself.

2.5 NSW government context

The following identifies some of the major policy and strategic NSW government drivers for open data.

2.5.1 NSW Government Open Data Policy

The NSW Open Data Policy was first issued in 2013 and refreshed in 2016.

[37] https://www.finance.nsw.gov.au/ict/sites/default/files/resources/NSW_Government_Open_Data_Policy_2016.pdf
[38] https://data.nsw.gov.au/data/organization
[39] https://www.finance.nsw.gov.au/ict/sites/default/files/resources/NSW_Government_Open_Data_Policy_2016.pdf, p5

The 2013 policy was developed as a deliverable under the 2012 NSW Government ICT Strategy. The ICT Strategy was a major initiative that 'elevated ICT to be one of the highest priorities of Government' and aimed to 'support the public sector to drive better service delivery, greater transparency and better value from investment in ICT'. [40]

The Strategy identified open data as a priority initiative. The Strategy sought to make government data available to industry and the community in order to stimulate the development of innovative approaches to service delivery. Key priorities under the NSW ICT Strategy relating to open data were to:

- Expand and enhance Data NSW
- Identify and publish syndicated feeds of high demand real-time data
- Prioritise key government datasets for standardisation in consultation with government, industry and the research community
- Define and agree on common data standards
- Develop and implement an open access licensing framework across government
- Host two apps4NSW events to leverage datasets
- Investigate opportunities to extend apps4nsw to include cross-jurisdictional datasets and collaboration. [41]

The 2013 Open Data Policy aimed 'to assist agencies across the NSW Government in embedding open data principles in their operations and in releasing high-value datasets' and 'to make appropriate government data available to industry and the community'. [42] The policy aimed to facilitate implementation of best practice open data principles across the NSW public sector and its purpose was to:

- Simplify and facilitate the release of appropriate data by NSW Government agencies
- Make explicit the NSW Government's commitment to open data and open government
- Create a practical policy framework that enables high-value datasets to be released to the public
- Help agencies in understanding community and industry priorities for open data
- Support the Government Information (Public Access) Act 2009 (NSW) (GIPAA) and promote simple and efficient compliance with the requirements set out in that Act.

[40] https://www.finance.nsw.gov.au/ict/sites/default/files/NSW%20Government%20ICT%20Strategy%202012_1.pdf, pp4-5
[41] https://www.finance.nsw.gov.au/ict/sites/default/files/NSW%20Government%20ICT%20Strategy%202012_1.pdf, p17
[42] https://www.finance.nsw.gov.au/ict/sites/default/files/resources/Open_Data_Policy.pdf

The policy had the community as a key focus and sought to drive value for the people of NSW by:

- Promoting open, transparent and accountable government
- Contributing to the digital economy in NSW, and promoting the development of new businesses and industries that can make use of government data
- Developing better public services
- Advancing citizen engagement with government and with the work of government
- Enabling data sharing between government agencies, in NSW and across jurisdictions
- Embedding open data principles across government operations
- Facilitating greater understanding amongst agencies of their own data and the potential of that data
- Supporting evidence-based policy making and policy research.

Following the development of the Data Sharing Act 2015 [43] and the establishment of the Data Analytics Centre, the Open Data Policy was refreshed in 2016. [44] It requires agencies to release data that is:

- open by default and protected when required
- prioritised, discoverable and useable
- primary and timely
- well managed, trusted and authoritative
- free when appropriate
- subject to public input.

The Open Data Action Plan [45] was issued in 2016 to drive implementation of the refreshed Open Data Policy. The Action Plan focussed on six key initiatives. The *Open Data Innovation Scorecard* initiative aimed to develop data maturity scorecards to guide agency progress towards open data. *Sustaining Open Data* aimed to develop a standard process to help agencies prioritise the development of APIs to share data. It also aimed to prioritise the release of high value data including fiscal, extraction industries, environmental and pollution and aggregated and statistical data. It sought to help agencies develop open data into their standard business approaches. The *dMarketplace* initiative aimed to deliver a

[43] http://www.legislation.nsw.gov.au/acts/2015-60.pdf
[44] https://www.finance.nsw.gov.au/ict/sites/default/files/resources/NSW_Government_Open_Data_Policy_2016.pdf
[45] https://www.finance.nsw.gov.au/ict/sites/default/files/resources/Open_Data_Action_Plan_2016.pdf

platform for industry and research to publish data connected with NSW and other government data. *Making Links with Data* aimed to improve the quality of data and metadata on Data NSW, to reduce the fragmentation and duplication of datasets and to help agencies apply safeguards to de-identify data prior to opening it. It also sought to prioritise publication of spatially enabled data in accessible formats. *Incentivising open data and fostering innovation* aimed to build partnerships with industry for investment to release data and pilot programs with universities to use data. It also sought to run annual open data awards and to publish data from research funded by government on Data NSW. *Connecting data and stories* aimed to establish an open data blog and engage the public in the design of datasets.

2.5.2 Government Information Public Access Act (GIPAA)

In 2009 the *Government Information (Public Access) Act* (GIPAA) established the NSW Government's commitment to opening government information, proactively and responsively, to the public. [46]

Section 3 of the Act specifically identifies that the object of the Act is to 'open government information to the public by authorising and encouraging the proactive public release of government information by agencies'.

Data NSW was created as a government portal in 2009 and supported the implementation of the GIPA Act by providing a platform for the proactive release of open government data.

The close relationship between the GIPA Act and open data is outlined in the NSW Open Data Policy 2013, the NSW Open Data Policy 2016 and throughout the website of the NSW Information and Privacy Commission, particularly the pages relating to Open Government [47] and the NSW Open Data Advocate. [48]

2.5.3 Open Data Advocate

In April 2016, with the launch of the NSW Open Data Policy in April 2016, the NSW Government appointed the Information Commissioner as NSW Open Data Advocate.

The Open Data Advocate's role is to encourage the release of data by NSW government agencies in ways that are respectful of data sharing safeguards. With the supporting structure of the *Government Information (Public Access) Act* 2009 (GIPA Act), the Open Data Advocate sees open data as having a significant impact in improving government, empowering citizens, creating opportunity and solving problems. [49]

To help to build an enabling environment in the NSW public sector that supports technical capability, governance and digital transformation, and also to build confidence in the community that the government is managing open data responsibly, the Open Data Advocate's current priorities are to:

[46] http://www.austlii.edu.au/au/legis/nsw/consol_act/giaa2009368/
[47] http://www.ipc.nsw.gov.au/open-government
[48] http://www.ipc.nsw.gov.au/open-data-advocate
[49] http://www.ipc.nsw.gov.au/open-data-advocate

- Guide agencies and citizens in promoting the availability of open access data
- Contribute to the development of indicators and measures to better recognise the most effective Open Data practices
- Promote Open Data as a cornerstone of Open Government. [50]

The Open Data Advocate actively promotes Data NSW [51] and is a key stakeholder DFSI's open data program and the Data NSW portal.

2.5.4 The Data Analytics Centre (DAC)

The DAC was opened in 2016 to facilitate data sharing between agencies and to manage whole-of-government analytics projects. [52] The objectives of the DAC are to provide insights into complex policy problems, support greater evidence-based decision-making and improve service delivery for the community.

The DAC works with data from a variety of sources, including open data, to derive insights that support strategic decision making, evidence based policy development and improved service delivery.

2.5.5 Open Government

The NSW Government is committed to the open government principles of transparency, participation, collaboration and innovation. Open data is a key enabler of open government initiatives.

The 2013 refresh of Data NSW was a deliverable of the NSW Government ICT Strategy [53] and represented the operationalisation of the government's commitment to open government, as established in the Premier's Memoranda M2012-10 *Open Government*.[54]

Making data more accessible was also a key component of the NSW Open Government commitment under the NSW ICT Strategy [55], and was a component the NSW Government's Digital Economy Industry Action Plan [56] which sought to implement open data innovation.

2.5.6 NSW data ecosystem

The NSW government is scoping the development of a data ecosystem environment. This environment is currently evolving but will likely encompass a range of data components.

[50] http://www.ipc.nsw.gov.au/sites/default/files/file_manager/Open%20Data%20Advocate%20Work%20Program%2016-17_0.pdf
[51] http://www.ipc.nsw.gov.au/open-data-advocate
[52] https://www.finance.nsw.gov.au/ict/nsw-data-analytics-centre
[53] NSW Government ICT Strategy 2012: https://www.finance.nsw.gov.au/ict/sites/default/files/NSW%20Government%20ICT%20Strategy%202012_1.pdf, pp15-17
[54] https://www.finance.nsw.gov.au/ict/resources/premiers-memorandum-m2012-10-open-government
[55] Open Government in NSW ICT Board Report (undated), https://www.finance.nsw.gov.au/ict/sites/default/files/ICT%20Board%20report%20on%20open%20government%20in%20NSW.pdf
[56] https://www.industry.nsw.gov.au/buy-from-nsw/industry-capabilities/information-and-communication-technology/industry-action-plan-digital-economy

Data NSW will need to be interoperable with the proposed ecosystem environment and supportive of business needs or data requirements from the ecosystem.

2.6 Scope and functionality of Data NSW

2.6.1 Overview

Data NSW is run by the DFSI.

Data NSW is a portal which means it is a gateway to NSW open data resources. March 2017 reporting identified 481 catalogue entries on Data NSW, providing access to over 100,000 datasets. [57] This included 2,728 datasets from Office of Environment and Heritage and 7,122 datasets from the Commonwealth government which are directly searchable from Data NSW. It also includes links to other NSW agency data portals such as HealthStats [58] with over 80,000 datasets.

Most of the datasets accessible via the Data NSW portal are hosted on sites run by NSW government agencies or clusters of agencies. This approach allows open data to be searched in a central Data NSW location but also made available via the custodian agency website in order to increase exposure and accessibility. The benefit of the central Data NSW portal is that there is one central location for searching and federation rather than a range of isolated agency, thematic or cluster-specific environments. Importantly however, individual agency, thematic or cluster approaches are not stifled. It increases transparency, accountability and innovation, while also stimulating the release of information and public participation.

Data NSW has also been federated with open data portals from other jurisdictions so that a search will provide links to datasets from South Australia, Queensland and the Commonwealth Government.

NSW Government has adopted an open access licensing framework to support the release and reuse of public information. Data that is listed on Data NSW has creative commons licencing. [59] Principally CC-BY licencing is applied, which allows unlimited reuse subject to attribution: 'the Creative Commons Attribution license allows re-distribution and re-use of a licensed work on the condition that the creator is appropriately credited'. [60]

Comparison of the March 2017 reporting with *ICT Metrics Report 2015 – 2016* (July 2016 reporting), shows that in this period, data accessible via the site increased by 37% (350 catalogue entries to 481 catalogue entries). Visitor numbers have also increased, from between 3,000 and 5,000 visits every month in the 2015 – 2016 reporting period to 5,500 – 7,000 visits per month in the March 2017 reporting period. Over time visitors have been

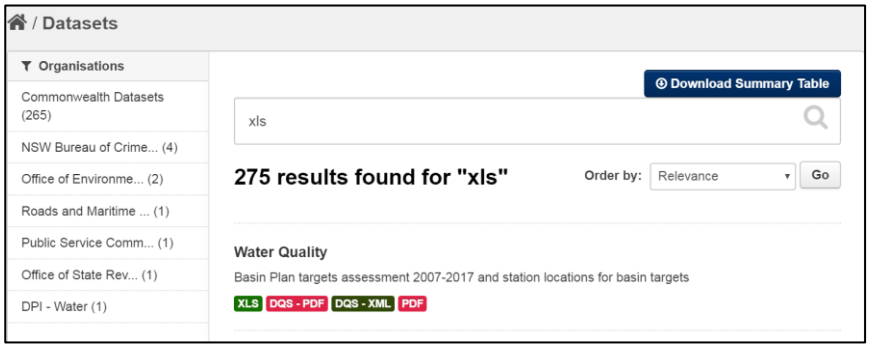
[57] NSW *State of the ICT Sector* reporting, 2017 (forthcoming)
[58] http://www.healthstats.nsw.gov.au/
[59] https://data.nsw.gov.au/copyright
[60] http://opendefinition.org/licenses/cc-by/

staying longer at Data NSW. The average number of page views per visit increased from 3.63 in July 2015 to 5.07 in July 2016. [61] The majority of visitors to the site are from Australia (89%), with the next largest group of visitors from the United States (3%).

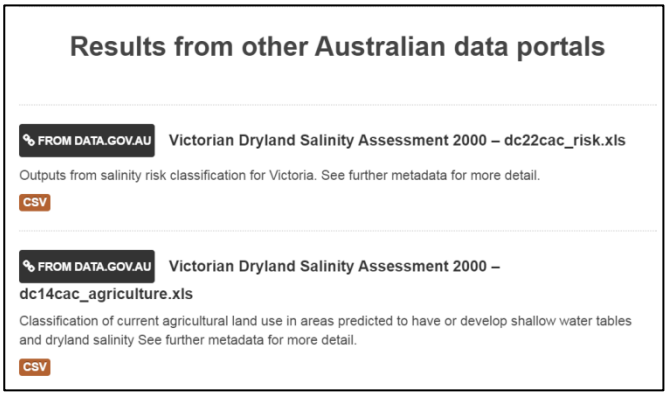
2.6.2 Key Data NSW functionality

a) Search and download datasets

The search function [62] enables datasets and related documentation to be searched and downloaded. Users can choose to view datasets, organisations and groups via the header menu, or search directly via the main search bar.

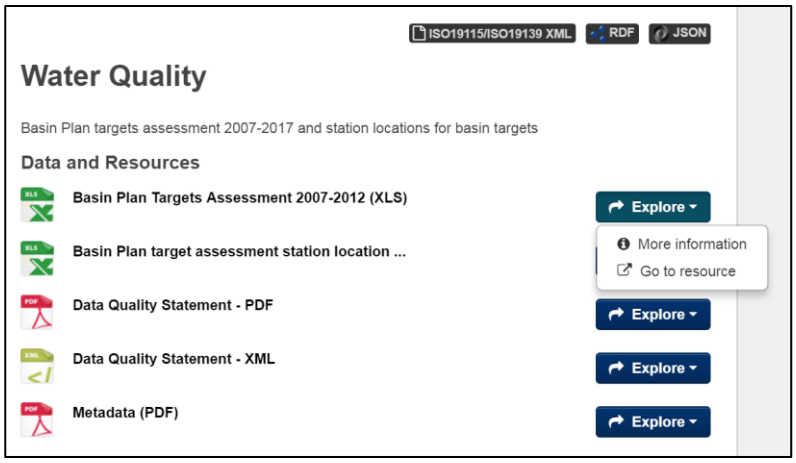


Results from searches show a summary preview with clear indications of the file types available. Results from other Australian data portals are displayed after NSW results and are labelled accordingly.



[61] ICT Metrics Report 2015-16, NSW Government. Available at: https://www.finance.nsw.gov.au/ict/priorities/ict-investment/ict-survey
[62] Accessible via https://data.nsw.gov.au/data

Having clicked the chosen link, the user is presented with the dataset. This may contain one or more files which can be viewed or downloaded.



Each dataset is licenced with a specific licence that describes how the data can be used. Most datasets published on Data NSW have a Creative Commons Share-Alike licence. This means that users are free to use the data provided and that they should openly share any data or products derived from the data use. Each dataset includes the metadata elements listed in the following table:

CKAN	Mandatory	Type	Standard
Title*	Yes	Free text	N/A
Description*	Yes	Free text	N/A
Licence*	Yes	Controlled list	Creative Commons 3, AusGOAL Restricted
Tags	No	Controlled list	APAIS Thesaurus
Groups (Function)*	Yes	Controlled list	Archives Investigator Function Set
Organisation*	Yes	Controlled list	Agency Name
Visibility*	Yes	Binary (yes/No)	N/A
Format*		Controlled list	Automatically Detected
Extra [Language]	No	Controlled list	Default to "English"
Extra [Publisher]	No	Controlled list	Agency Name
Extra [Type]*	Yes	Controlled list	DCMI Type Vocabulary
URL*	Yes	Free text	Link to source file only e.g. example.nsw.gov.au/files/foo.xls
[date added, date modified]*	Yes	N/A	Automatically generated

* mandatory

This metadata is intended to help users understand datasets and the context they were produced in, and determine whether they are fit for purpose.

2.6.3 Data NSW roles and responsibilities

a) User roles

Agencies play an active role in uploading data resource information (metadata) to Data NSW. Each agency, therefore, must have its own user account and login details. To apply for a user account, users must complete and email a user account creation form and email to DFSI.

Agencies are assigned one of two levels of access to data.nsw.gov:

- Data Editor – have the ability to upload datasets to their agency's section of the website
- Trusted Data Editor – can upload and publish dataset information live to Data NSW. Agency users will need to meet certain criteria to gain Trusted Data Editor status. These include undergoing Data NSW training and having demonstrated ability to use the Data NSW website to professional standards. Trusted Data Editors are assigned on a case-by-case basis, taking into consideration the needs and specific requirements of each agency.

Nominated DFSI staff also take on site administrator, editor and publisher roles.

A summary table below sets out the roles and access rights of the different user roles on the Data NSW website.

Role Type	Users	Upload	Publish	Admin Roles
Administrator	DFS Staff	Yes	Yes	Yes
Editor	DFS Staff	Yes	Yes	Partial
Trusted data editor	Agency users (selelected)	Yes	Yes	No
Data editor	Agency users (all)	Yes	No	No
Publisher	DFS Staff	No	Yes	No
Observer	All	No	No	No

A register is maintained that outlines all Data NSW roles and the names and contact details of the specific agency and DFSI staff members filling these roles.

b) Custodianship responsibilities

Data NSW does not store data, but links to datasets stored on agency websites. Data NSW is also not responsible for the storage or availability of datasets linked on Data NSW website. Agencies maintain responsibility for the linked dataset and the metadata they have provided. Agencies have the ability to modify their datasets and metadata and are encouraged to correct oversights or previous inaccuracies. The uploading agency also has the responsibility to ensure the data has been authorised for public release. Data

custodianship remains with the publishing agency, as described in the NSW Data and Information Custodianship Policy. [63]

c) Data quality statements

Data quality statements or other additional information about a dataset that agencies may wish to provide can be uploaded as a data resource within the dataset.

DFSI has developed a data quality tool that agencies are able to use to generate data quality statements. [64] This tool allows agencies to generate descriptive statements about their datasets that will enable users to effectively understand and make use of these datasets.

As identified in the NSW ICT Metrics Report 2015-16, datasets that include a data quality statement are downloaded almost twice as much as datasets without. [65] This highlights the importance of publishing data with the appropriate metadata to enable its usability. While the proportion of machine readable datasets and datasets with data quality statement are increasing on Data NSW, they are currently low. There is an opportunity to build on this foundation to significantly improve the value of open data by improving readability and interpretability.

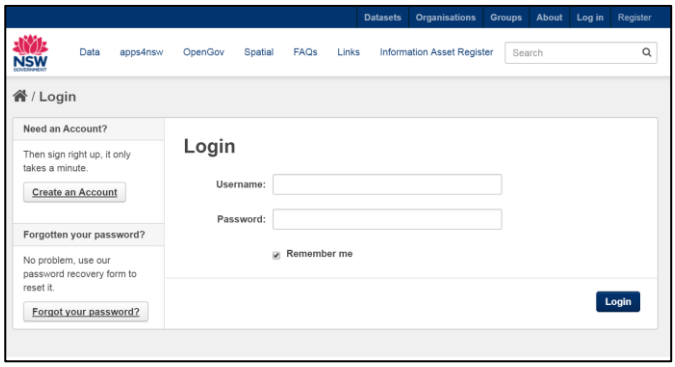
The following table identifies the proportion of datasets available on Data NSW with data quality statements, as well as machine processable data and spatially-enabled data. [66]

	1 June 2015	*1 June 2016*
Data quality statements	0.4%	6.1%
Tabular (machine processable data)	22%	27%
Spatial	2.4%	3.0%

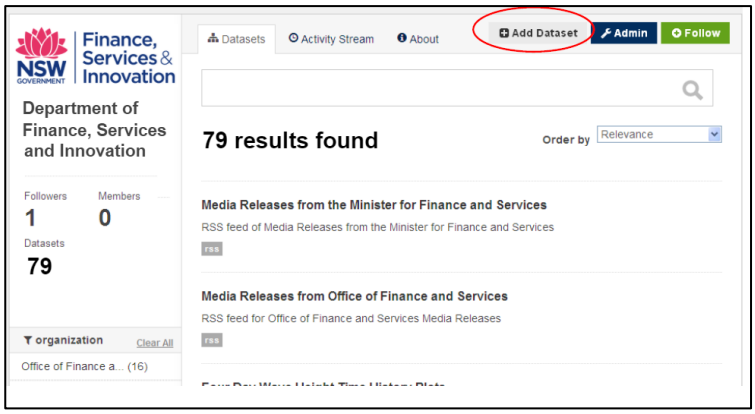
d) Responsibilities for loading datasets to Data NSW

To load datasets, agency staff login using their agency account details:

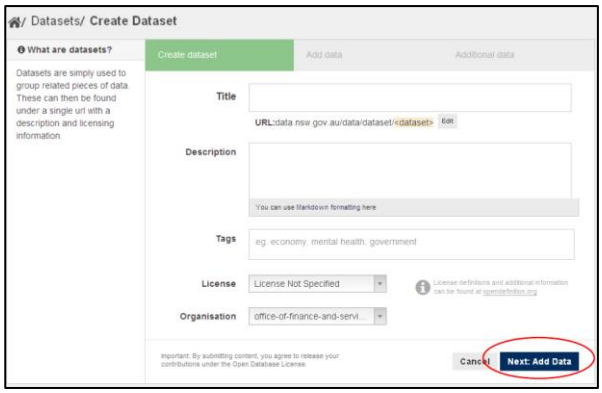
[63] https://www.finance.nsw.gov.au/ict/resources/nsw-government-data-and-information-custodianship-policy
[64] This tool is in beta testing and is available at http://web.fnsw.links.com.au/identify
[65] https://www.finance.nsw.gov.au/ict/sites/default/files/resources/150721_ICT%20Metrics%20Program_REPORT_V6b.pdf, p26
[66] The data from this table was reported in the NW ICT Metrics Survey Report, 2015-16, https://www.finance.nsw.gov.au/ict/sites/default/files/resources/150721_ICT%20Metrics%20Program_REPORT_V6b.pdf, p26



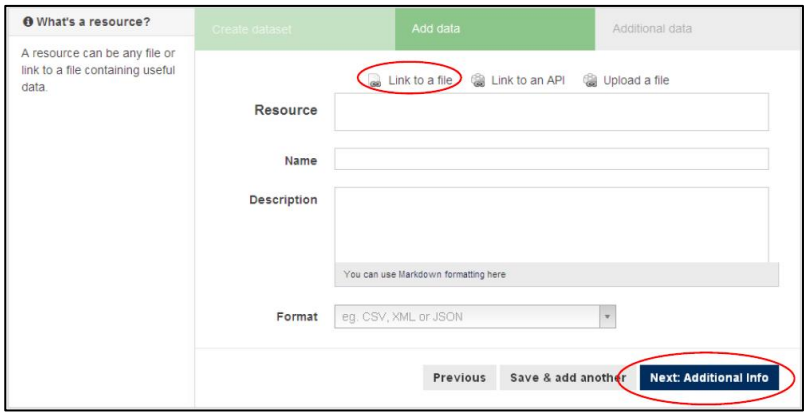
They navigate to their agency's Organisation page, either by clicking on "Organisations" on the top navigation pane and selecting their organisation's name, or by navigating to data.nsw.gov.au/your-agency-name. They can then click on "Add Dataset".



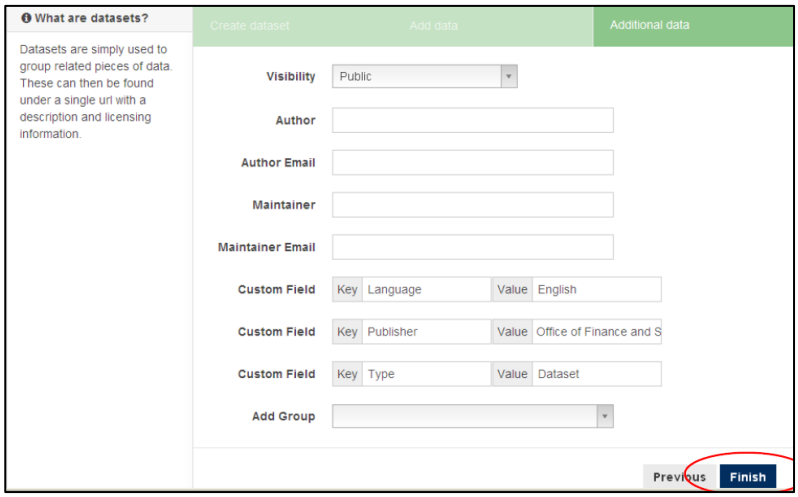
Agencies then fill in all required metadata to describe their dataset.



Then agencies create a link to a data file or API on their website.



The final step is for agencies to add custodian and contact details for the dataset.



Clicking Finish will upload the metadata description and data link to the agency's data collection on Data NSW.

e) Other features of Data NSW

Social Media

The front page displays recent *Twitter* activity (posts and reposts only) on behalf of @datansw. An open data blog hosts posts relating open data in NSW.

Data NSW blog

The blog area [67] of Data NSW posts stories, case studies and updates about Data NSW.

Request a dataset

Data NSW provides a 'request a dataset' button for users to click to request a datasets be made available. These requests are automatically routed to the central @datansw inbox for action.

[67] See https://data.nsw.gov.au/blog

The Information Asset Register (IAR)

The purpose of the IAR [68] is to facilitate data sharing between agencies.

The IAR provides searchable metadata and contact details for a list of core-value information assets, including datasets, held by NSW government agencies. The IAR does not host data, only metadata about datasets.

The information in these datasets is more sensitive than those made available via Data NSW. Only registered NSW Government employees are able to access the IAR. Metadata contained in the IAR and any information from the datasets listed is only for public release with the express permission of its custodian.

Apps Showcase

The Data NSW Apps Showcase [69] is an information page that provides details of apps that have been created using openly available data, or data that has been made available to entrants in competitions (such as apps4NSW) or hackathons.

apps4nsw

The apps4nsw information page [70] describes the apps4NSW program of events that aimed to encourage the use of NSW Government open data to create innovative web and mobile applications. The apps4NSW competitions ran between 2010 and 2015.

Open Data Action Plan

Using traffic light indicators, the Open Data Action Plan page [71] provides a visual depiction of DFSI progress against the Open Data Action Plan [72] and its deliverables.

Action	Jun 2016	Sep 2016	Dec 2016	Jun 2017	Dec 2017
Develop a data request service so that the public can easily request data to be released or updated	✓				

f) Technical components of Data NSW

Technically, Data NSW is composed of a variety of components:

- SKU: CKAN-ODM (Medium On-Demand CKAN Open Data Solution, with all services provisioned within Amazon Web Services in the Sydney region)
- Drupal content management system that supports the blog functionality

In March 2017 Data NSW moved to a secure https URL format.

[68] See https://data.nsw.gov.au/iar
[69] See https://data.nsw.gov.au/apps
[70] See https://data.nsw.gov.au/apps4nsw
[71] See https://data.nsw.gov.au/plan
[72] https://www.finance.nsw.gov.au/ict/resources/open-data-action-plan

2.6.4 Program logic

The program logic for Data NSW is shown in the figure below. This figure shows how inputs and activities are related to the program outcomes.

2.7

- assess the extent to which Data NSW is being implemented as intended;
- determine the extent to which the objectives of Data NSW are being met;
- determine if Data NSW has contributed to improved outcomes for users;
- and assess the cost-effectiveness of the Data NSW.

The method is outlined in the next chapter.

3. Evaluation methodology

This section of the report sets out the methodology that was used to conduct the evaluation of Data NSW.

3.1 Evaluation Scope

The evaluation sought to answer eight key questions mapped against three domains as outlined in the table below. The methods and data sources are outlined in more detail in subsequent sections.

Domains	Evaluation questions
Implementation: Appropriateness	• To what extent does Data NSW address an identified need? • How well does Data NSW align with government and agency priorities?
Implementation: Efficiency	• Is Data NSW being implemented appropriately? • Are users being reached as intended? • How satisfied are users? • Has Data NSW delivered value for money?
Outcomes: Effectiveness	• To what extent has the Data NSW achieved its objectives? • Were there any unintended impacts (positive or negative)?

3.2 Evaluation methods

The following methods informed the evaluation:

- A review of guidelines and policies supporting and informing Data NSW
- A review of broader relevant NSW guidelines, policies and initiatives
- Stakeholder consultation with the DFSI developers of Data NSW
- A survey of Data NSW users, stakeholders and contributors in the NSW public sector, and users in the private sector
- A high level literature scan to identify comparable models within Australia and internationally
- A detailed assessment of comparable open data portals

- An assessment of metrics, statistics and reporting compiled by the Data NSW reporting tools
- A review of international reports, reviews and evaluations on the effectiveness and sustainability of open data portals.

3.3 Evaluation strengths and limitations

Limitations of the evaluation primarily relate to:

- Limited ongoing administrative data collected from the inception of Data NSW
- Absence of baseline performance data prior to the implementation of Data NSW
- Completeness and accuracy of data
- Timeframes for effective stakeholder consultation.

Strengths of the evaluation are that it was:

- Conducted by business staff actively engaged with the Data NSW portal
- Performed when significant international reviews of government open data portals are underway and so was able to draw from major qualitative and quantitative studies from Australia and overseas
- Able to utliise a wide range of sources to identify a broad range of lessons.

4. Key findings: Implementation – Appropriateness

4.1 To what extent does Data NSW address an identified need?

4.1.1 Comparison to national and international data portal benchmarks

In the following assessments the functionality of a range of data portals is assessed and compared to the functionality of Data NSW, to help address the question of whether Data NSW addresses an identified need.

The following portals have been selected as they:

- Represent national and international portal models at federal, state and local government levels, and in the private sector
- Are primarily portals from jurisdictions that rank highly in the Open Data Barometer [73], an annual report that ranks governments on their readiness for open data, the implementation of open data programs and the impact that open data is having on their business, politics and society

a) Comparability to the Commonwealth Government's open data portal

Background

With 537 local governments, eight state and territory governments and one federal government, releasing open government data through a single portal is a considerable challenge for Australia. The Australian Commonwealth Government aims to overcome this challenge with its open data portal, data.gov.au. The portal was first launched in 2009 following the Government's Declaration of Open Government and as a response to the Government 2.0 Taskforce Report.[74] It's run by the Commonwealth Department of Finance.

The vision of data.gov.au is to make all government data in Australia publically discoverable in a single place, machine readable, API enabled, permissibly licensed with some visualisation functionality.[75]

[73] http://opendatabarometer.org/
[74] About data.gov.au, Australia Government, Available at: https://data.gov.au/about
[75] Open Data Toolkit, Australian Government, Available at: https://toolkit.data.gov.au/index.php?title=Introduction#What_is_data.gov.au

The platform has undergone a number of improvements since its first went live in 2009. Most notably, in July 2013, data.gov.au was upgraded as a CKAN platform – the same platform that hosts Data NSW. The upgrade also included the introduction of spatial data hosting.[76]

At the time of this evaluation, the platform provided access to[77]:

- 49.9K discoverable datasets (compared to over 84K available through Data NSW)
- 9052 machine readable/data API resources
- Data files/resources from 519 organisations
- 9K API enabled resources
- 36.9k openly licensed datasets
- 24 unpublished datasets

The number of datasets available on data.gov.au increased markedly from 5,749 in 6 February 2017 to 47,806 in May 22 2017[78]. The reason for the spike in numbers is unknown to the evaluators however it is predicted that this is due to the improved search functionality being implemented (discussed in further detail below).

As part of the current National Innovation and Science Agenda, the CSIRO's data and digital research group, Data61, is currently working with the Department of Prime Minister and Cabinet to develop the next generation of data.gov.au. Key features will include:

- Next generation search discovery capability.[79]
- Data collaboration tools to connect people, tools, services and open data.[80]
- Data cards to provide discoverable insights and data for the public.[81]

Further details are on each of these features are provided in the sub-sections below.

Improved search discovery

A key feature of the next generation of the portal includes the addition of a unified search portal (http://search.data.gov.au/)**.** The intention of the new search interface is to enable all government open data released in Australia to be discoverable in a single place. The new approach will result in improved discoverability of open data available in Australia. In

[76] Have your say on the next generation of data.gov.au, blog.data.gov.au, Available at: https://blog.data.gov.au/news-media/blog/have-your-say-next-generation-datagovau
[77] Data.gov.au home page, Australian Government, Available at: http://data.gov.au/
[78] Total number of Datasets, data.gov.au, Available at: http://data.gov.au/stats#total-datasets
[79] Next generation search and discovery, data.gov.au technology preview, Available at: http://preview.data.gov.au/search.html
[80] Why Data Collaboration? Data.gov.au technology preview. Available at: http://preview.data.gov.au/data_collab/01.html
[81] Discoverable insights and data for everyone. Data.gov.au technology preview. Available at: http://preview.data.gov.au/data_cards/01.html

addition, improvements will be made to the presentation of search results and filtering tools.

The new search interface has built-in connectors to the following CKAN-based open data government portals in Australia:

- https://data.gov.au
- http://data.nsw.gov.au
- https://www.data.vic.gov.au
- https://data.qld.gov.au
- https://data.sa.gov.au
- http://catalogue.beta.data.wa.gov.au
- https://www.data.brisbane.qld.gov.au

The interface also has CSW-based connectors for the following services:

- http://www.ga.gov.au/geonetwork/srv/en/
- http://www.bom.gov.au/geonetwork/srv/eng/
- http://data.aims.gov.au/geonetwork/srv/eng/
- http://catalogue.aodn.org.au/geonetwork/srv/eng/
- http://www.marlin.csiro.au/geonetwork/srv/eng/
- http://data.auscover.org.au/geonetwork/srv/eng/
- http://www.mrt.tas.gov.au/web-catalogue/srv/eng/
- https://data.thelist.tas.gov.au:443/datagn/srv/eng/
- https://sdi.nsw.gov.au/

These connectors will regularly source metadata from the above portals into a common index. Data61 is also exploring other connects to enable connections to Socrata and ArcGIS Open Data based portals. This added functionality will be more efficient compared to what was previously available through CKAN. The Commonwealth Government will extend access to this API-enabled search functionality to other open data portals across Australia. This will ensure that other open data portals also have augmented searching.

Data collaboration spaces

The new generation of the portal will include an API-enabled platform that connects new and existing tools and services with data and people. The intention of the platform is to improve the quality, range and depth of open data. The tool will connect data publishers

with data consumers so that errors in publically available data can be better identified and corrected for, and analysis can be reviewed and refined before it is released.

Data cards

The portal will include the addition of short, easy to read snippets of data-driven information, referred to as data cards. The intention of this feature is to enable improved discovery of data insights and its underlying data. It will provide a central platform for users to discover the source of data, methods used for data analysis, and insights and outputs related the data. This feature is being designed to enable citizens to track and better understand the source of open data findings presented by the media, government, social media channels, advocacy groups, and commercial organisations.

Comparative overview

There are a number of features that Data NSW shares in common with data.gov.au. This includes:

- A home page that includes features other than the open data catalogue.
- An open data blog.
- Access to apps developed from open data.
- Users are able to subscribe to new dataset feeds or customise feeds that they are interested in.
- Both sites provide links to websites containing map-based spatial data. The website providing spatial data from Commonwealth government agencies is NationalMap. In NSW the comparative website is NSW Globe.

Functionality unique to Data NSW includes:

- The Information Asset Register.
- A twitter feed.
- Easily visible contact email address for users to query issues or to request datasets not available through the Open Data Portal. In contrast contact details aren't easy to locate on data.gov.au and users aren't invited to use the email address to query issues. However, the data.gov.au provides a function for allowing users to submit a request for data and any requests are publically displayed on the website. In comparison Data NSW does not publish data requests.
- Links to the NSW Open Data Policy.

Functionality unique to data.gov.au that should be considered as part of future upgrades to Data NSW is set out in the section below.

Learnings

There are a number of features of data.gov.au that should be considered to improve Data NSW. These include the:

1) **look and feel of the home page.** The central feature of the data.gov.au home page is a search tool (see screenshot below taken from http://data.gov.au/). In contrast data NSW does not include a search bar for data on the homepage (see second screenshot below). Data NSW should consider adding a clearly visible search tool function on its home page given that the primary reason that users visit the site is to access the open data catalogue. In addition the data.gov.au home page presents less script, which provides an enhanced user orientation experience in comparison to Data NSW.

Figure 1: Screenshot of the data.gov.au home page

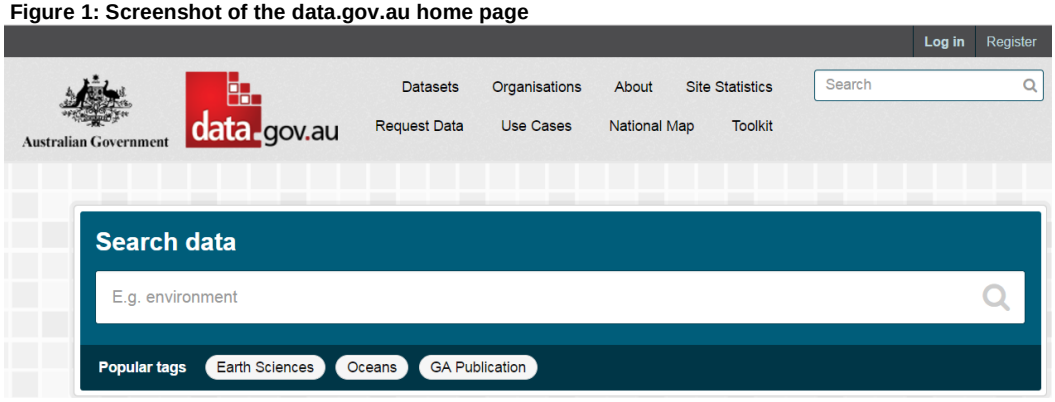
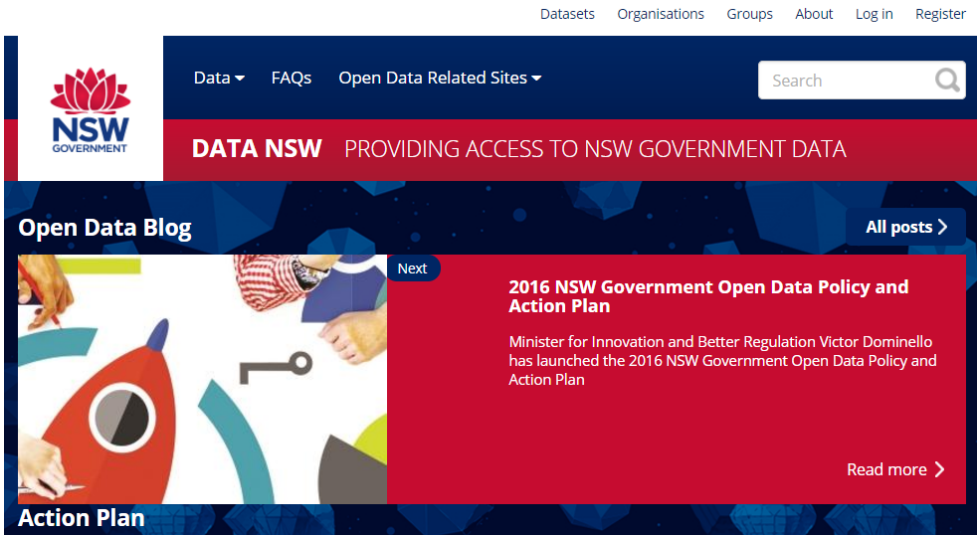


Figure 2: Screenshot of the Data NSW home page

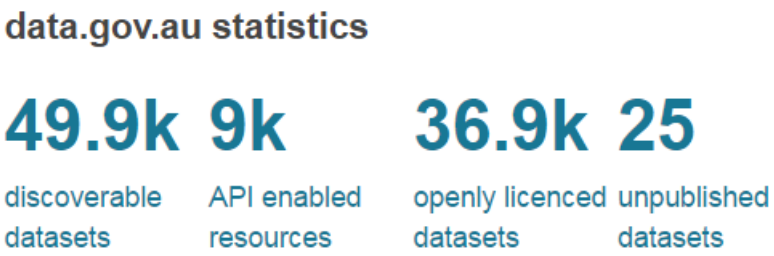


2) **new functionality of data.gov.au (including the improved search functionality, data cards, and collaboration spaces).** The new functionality of the next generation of data.gov.au should be evaluated and considered as part of any updates to Data NSW. In particular, NSW Government should work with the Commonwealth Government to gain access to the new API-enabled search functionality to augment the current search functionality of Data NSW.

3) **open data toolkit.** The data.gov.au home page provides a link to an open data toolkit for agencies. The toolkit provides information on how agencies can make data open, the benefits of open data, technical and other support related instructions for using data.gov.au, and advice on policy developed and planning requirements for creating datasets and making data open. In comparison the Data NSW home page instructs users to contact the Data NSW mailbox for assistance in publishing agency data. A key recommendation of this review is that the NSW Government should also develop a toolkit to consolidate its existing open data guidance material and to provide agencies with clear steps on how to publish data on Data NSW. This will minimise unnecessary queries to the Data NSW website.

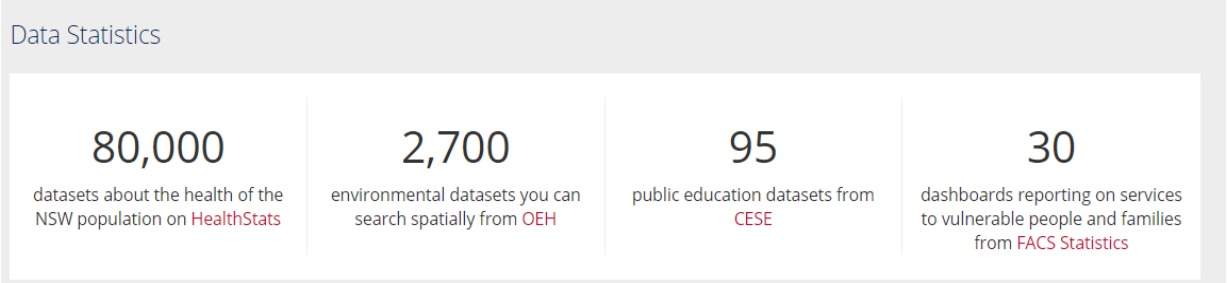
4) **Site statistics.** One of the most visible features of the data.gov.au website is the statistics on its open data catalogue (see Figure 4 below). Statistics are provided on the number of discoverable datasets, API enabled resources, openly licenses datasets and unpublished datasets.

Figure 3: Screenshot of the statistics displayed on the data.gov.au home page



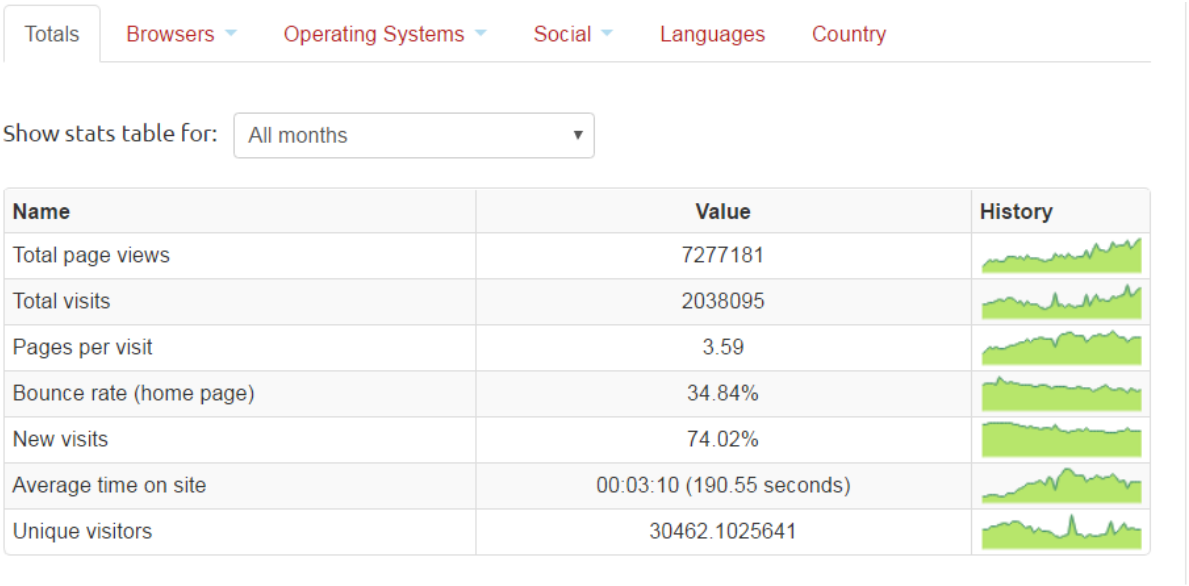
Although Data NSW homepage provides statistics on the number of datasets published by open data platforms in NSW (including HealthStats, the Office of Environment and Heritage, the Centre of Evaluation and Statistics and Family and Community Services Statistics), it does not provide a consolidated number of datasets that are publically available through the website (see Figure 5 below). The additional statistics reported on the data.gov.au homepage should also be considered for inclusion on the Data NSW homepage.

Figure 4: Screenshot of the statistics displayed on the Data NSW home page



In addition, data.gov.au provides detailed, real time statistics on the number and type of visits to the site (see Figure 5). This feature should also be considered for future upgrades to Data NSW.

Figure 5: Screen shot of the site statistics reported on the data.gov.au website



b) The World Bank: Databank

Background

In 2010, as part of its mission to work towards a world free of poverty and to provide financial and technical assistance to developing countries, the World Bank began opening its data via DataBank. [82]. The World Bank's view is that 'timely and reliable statistics are key inputs to the broad development strategy. Improvements in the quality and quantity of data on all aspects of development are essential if we are to achieve the goal of a world without poverty'. [83] They have therefore invested heavily in opening their global information assets as open data.

DataBank makes WorldBank databases open to all users, providing access to more than 7000 financial, business, health, economic, and human development indicators. A data catalogue provides access to databases including those containing the World Development Indicators, Africa Development Indicators, Global Economic Monitor and Global Development Finance. DataBank offers access to data from more than two hundred economies, with many time series going back fifty years. Entire datasets for country or indicators can be downloaded, raw data can be accessed and users can comment and ask questions about data. [84]

Other key features include:

- Quick access to data from more than 55 databases

[82] databank.worldbank.org
[83] http://data.worldbank.org/about
[84] https://developmentseed.org/blog/2010/apr/20/world-bank-open-data-initiative-launched-on-drupal/

- Enhanced visualizations and reporting capabilities
- APIs for programmatic access to data and metadata
- Dashboards for each country with key metrics
- Data and metadata (definitions and source notes) downloads in multiple formats
- Ability to create and share customized reports including tables, charts and maps
- Mapping capacity, including sub-national mapping for selected databases
- Availability of indicators in five languages – English, Spanish, French, Arabic and Chinese
- Widgets to embed data graphs or visualisations on blogs and web sites
- Social media integration
- Tablet compatibility
- A case management system to allow users to track the status of their data requests
- A comprehensive suite of documents and toolkits to support data publication

Previously access to this information was available only to paying subscribers. But by removing the pay wall and releasing their data, the Bank wanted to 'encourage innovative analysis of development issues and problems and stimulate evidence-based policy making in developing countries' and to support the efforts of national data-producing institutions. [85] The Bank also chose to release its data in order to position itself as a transparency leader, and to enable broader multi-national participation in the Bank's operations. The approach was immediately successful. By 2012 DataBank was averaging 740,000 visits per month. Between 2011 and 2012 the DataBank query tool received 2.8 million visits and DataBank's API was used by about 250,000 unique IPs. [86] Now one third of all web traffic at the World Bank is for open data. [87]

Comparative overview

DataBank offers substantially more dataset interactivity than Data NSW and users are able to customise, view, report and analyse specific data components much more thoroughly through the DataBank site.

This is because the World Bank has invested significantly in data consistency and quality. Much of the data provided through DataBank comes from the statistical systems of member countries and the quality of DataBank data is dependent on the effectiveness of these national systems. The World Bank therefore works to develop the capacity of

[85] http://siteresources.worldbank.org/EXTANNREP2010/Resources/WorldBank-AnnualReport2010.pdf, 12
[86] http://blogs.worldbank.org/opendata/whats-the-most-popular-world-bank-open-data
[87] http://opendatatoolkit.worldbank.org/docs/opendatatechnicaltrainingeap.pdf

national statistical systems. Their view is that without better and more comprehensive national data, it is impossible to develop effective policies, monitor the implementation of poverty reduction strategies, or monitor progress towards global goals, and therefore it invests heavily in data consistency and quality. [88]

It has particularly invested in the development of statistical indictors, data elements that represent data for a specified time, place or other characteristics. For example, within the broader indicator Economy & Growth there are dozens of sub indicators like GDP Growth (annual %), GDP per capita (current US$), Gross Capital Formation (% of GDP), Gross Savings (% of GDP). When Data Bank was launched in 2010 it enabled access to 2000 indicators. By investing in data capacity and quality across member nations, DataBank now provides access to more than 7000 indicators. [89] The WorldBank has also invested significantly in APIs to make its data available and useable.

Much of the World Bank's data-related activities have focussed on improvements in the quality and quantity of data. The World Bank:

- Invests in data forums to develop appropriate frameworks, guidance and standards of good practice
- Works to build consensus and define internationally agreed indicators
- Implements professional standards for the collection, compilation and dissemination of data to ensure confidence in data quality and integrity
- Promotes standard practices and methodologies to ensure data consistency across countries
- Establishes data exchange processes and methods
- Invests in spatially enabling all project data, all relevant indicators and any other data that needs to be displayed on a map
- Provides training in data practices to increase the accessibility and availability of data knowledge
- Responds to large volumes of queries and requests from inside and outside the Bank
- Works to maintain the relevance of data, by ensuring data collected continues to meet the needs of users and policymakers

The World Bank is also a consumer and communicator about its data. A 2014 evaluation of DataBank found that some datasets were not being downloaded. As result, the World Bank began a new proactive media-led approach to data which repackaged information in more accessible ways. According to Tariq Khokhar, Data Scientist and the institution's

[88] http://data.worldbank.org/about
[89] http://data.worldbank.org/indicator

Global Data Editor, communication, staff were trained to make eye catching charts and headlines such as 'Most Refugees don't live in camps', thousands of such charts and headlines have since been created with simple chart making tools. Datasets are used to generate and share global insights. [90] This media-led approach has also led to a significant social media presence and substantial community engagement with DataBank data through these channels. [91]

Learnings

Much of the World Bank's success has come from its substantial investment in data consistency and quality. By building standardisation and by focussing on the data maturity in countries around the world that need to supply it with data, the World Bank has been able to build an exceptionally large, useable, customisable and adaptable collection of data for global consumption.

To build such useable and consistent data, Data NSW could:

- Work with key data producers in certain business areas to define standard practices and methodologies
- Publish these standard practices and methodologies to help build further consistency
- Consider publishing indicators for certain areas of business
- Work with data producers to build spatial enablement into relevant datasets
- Promote and advertise data releases in ways that engage the community and that share data-driven insights with the widest possible audience.

c) UK Government – data.uk.gov

Background

The United Kingdom government is recognised as the world leader in open data. Recent research supports this conclusion. A 2017 evaluation of sixty-seven national data portals rated the UK's data portal the highest among the evaluated countries. [92] The UK is also ranked number one in the world in the Open Data Barometer (ODB) Report, an international research initiative that assesses national open data readiness, implementation and impact in countries across the world. The ODB Report has ranked the

[90] http://blogs.worldbank.org/voices/year-review-2016-12-charts-and-video
[91] See for instance the World Bank's very active Twitter account, https://twitter.com/worldbankdata?lang=en
[92] Renata Máchová, & M. Lnenicka, Evaluating the Quality of Open Data Portals on the National Level, *Journal of theoretical and applied electronic commerce research*, Volume 12 No 1, 2017, accessed via http://dx.doi.org/10.4067/S0718-18762017000100003

United Kingdom as the leading open data nation in the world each year since the report's inception in 2014. [93]

The key publicly stated drivers for data.gov.au are community focussed. The site states that 'government is releasing public data to become more transparent and foster innovation' and to make it 'easier for people to make decisions and suggestions about government policies based on detailed information'. [94]

Data.gov.uk was launched in closed beta in September 2009 and publicly launched in January 2010. Support for data.gov.uk continues to be strong. In February 2015 the site contained over 19,343 datasets from a wide range of UK Government departments. By May 2017 this figure had grown to over 43,000 accessible datasets. Driving this growth is the fact that the UK has established hard targets for data release to drive ongoing population of data.gov.uk. Ministerial Letters to government departments have been used to set targets for the public release of specific datasets, and agencies have also been directed to open specific numbers of datasets. For example, in June 2015 the Department of Environment, Food and Rural Affairs was directed by its Secretary of State to open 8000 datasets within 18 months. [95] It ultimately released 11,000 datasets by June 2016. [96]

Public use of data.gov.uk is also strong. Reporting statistics from September 2016 show that:

- The site has 162,486 (0.25% of inhabitants) unique visitors on average per month
- 19.4% of visitors are from beyond the UK
- 67% of traffic is generated by humans
- Most data sets are available in CSV format
- 50-60% of datasets on the site are machine readable. [97]

Public interface

Data.gov.uk offers a clean and simple interface for users to search and retrieve data. Up until 2016 the data.gov.uk homepage highlighted open data activity. However research into user needs demonstrated that users' key interest is in data and so the page was redesigned in 2016 to make data and data searching its focus. [98]

Data.gov.uk runs on a CKAN platform and offers standard CKAN search features that allow users to view the entire data catalogue or find data using keyword, category, theme, file format or creating department search options. Each dataset displays advice for

[93] The United Kingdom has been rated Number 1 in each of the four annual editions of the Open Data Barometer since 2014. See http://opendatabarometer.org/
[94] https://data.gov.uk/about
[95] http://www.ipc.nsw.gov.au/sites/default/files/file_manager/Conditions_Enabling_Open_Data_Report_Final.pdf, p7
[96] https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf, pp63-64
[97] Statistics available from https://www.europeandataportal.eu/sites/default/files/country-factsheet_united-kingdom.pdf
[98] https://data.blog.gov.uk/2016/08/05/updating-the-data-gov-uk-homepage/

accessing and interpreting its data. [99] Data.gov.uk is also currently trialling APIs for releasing data.gov.uk datasets. [100] The CKAN platform that that data.gov.uk is built on also has an API that allows developers to perform nearly any action that's also available to the users of data.gov.uk, meaning the API can be used to:

- show metadata about a single dataset;
- find publishers and the datasets they publish;
- search for all CSV files published by a specific publisher, within a date range, using search keywords;
- publish metadata, if you have the appropriate permissions. [101]

The CKAN framework means that any functionality added to the user interface automatically makes it available as an API. This promotes designing, building and using APIs as a critical component of developing the system. [102]

Other key public interface features of data.gov.uk include:

- Public rating of datasets
- Public comments on datasets
- Embedded maps showing the location datasets relate to [103]
- Ability to subscribe to updates when datasets are updated
- Ability to email the dataset owner about dataset questions
- Use of the Open Government License, a variant of the Creative Commons Approach that is more adaptable to Crown copyright issues. [104]
- Ability to requests new datasets and functionality that publicly tracks and provides a status update on all current requests [105]
- An apps area that lists 412 apps built with public data and provides users with the functionality to upload their own apps or visualisations, and ability to comment on apps and visualisations. [106]

[99] See for example https://data.gov.uk/dataset/conservation-areas
[100] https://data.gov.uk/data/api/
[101] See https://data.blog.gov.uk/2016/08/23/api-first-at-data-gov-uk/
[102] For a full description of the CKAN API see http://docs.ckan.org/en/latest/api/index.html
[103] https://data.gov.uk/location/map-based-search
[104] More information about the Open Government Licence is available at http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/
[105] https://data.gov.uk/data-request
[106] https://data.gov.uk/apps

Social media and user engagement

Data.gov.uk is supported by an active social media presence. The Twitter account @DataGovUk promotes datasets, open data initiatives and the broad use of data. [107] This account is also active in responding to open data enquiries. The *Data in Government Blog* is also regularly updated with updates on projects and research relating to open data. [108] A forum was also hosted for several years on data.gov.uk to foster data-related discussion and an open government community. [109]

Comparative overview

The success of data.gov.uk is attributable to:

- Strong governmental drivers for open data
- A Government Data Programme
- A strategic and coordinated approach to data and data quality
- A dual focus on public and public sector user needs

Strong governmental drivers and enablers for open data

The UK has a strong enabling environment that drives its open data culture. Each Department is required to develop and publish an Open Data Strategy, indicating what data they are committing to release, and by when. The Public Sector Transparency Board monitors these strategies and assesses progress against them. If departments do not meet their data publication targets, they are called before the Board to discuss and implement mechanisms that will enable the department to meet its open data targets.

There is detailed and consistent guidance in place on anonymisation and de-identification which provides practical case studies, examples and detailed advice for agencies to enable open data release. [110]

A Government Data Programme

The UK's Government Data Programme works to:

- Enable better operational, policy and economic decisions in government, dramatically cutting costs and improving the way in which government works as a result of modern data access and data science

[107] https://twitter.com/DataGovUK
[108] https://data.blog.gov.uk/
[109] See https://data.gov.uk/forum
[110] See Information Commissioner Office's Anonymisation: Managing Data Protection Risk 2012 and UKAN Members (includes members from Open Data Institute, Information Commissioner's Office, University of Oxford, London School of Economics, National Archives and others). Mark Elliot, Elaine Mackey, Keiron O'Hara and Caroline Tudor, *The Anonymisation Decision-Making Framework* (July 2016).

- Create the next wave of digital platforms on which we can build services for the benefit of citizens, underpinned by a new data infrastructure across government
- Boost to the UK economy as a result of better quality data opened up to those building data businesses
- Retain citizen trust with a clear approach about what should and should not be done with these powerful tools. [111]

To achieve these aims, the Programme focusses on three key areas: making better use of data, creating a modern data infrastructure, and through common data policies and governance. [112]

A strategic and coordinated approach to data and data quality

The Shakespeare Review of Public Sector Information [113] found that government needed to improve the quality and availability of public data. The National Information Infrastructure has been developed in response to:

- Make the most critical government data easier to find in one place
- Improve the quality and interoperability of the data through clarity about data standards
- Pay attention to the usability of the data, with API access alongside bulk download, and service level agreements about the persistence and timeliness of the data. [114]

The National Information Infrastructure also identifies core national datasets. To ensure the appropriate management of these datasets it also includes:

- Guiding principles
- A governance structure
- Baseline quality criteria
- Vocabularies and code lists
- Licensing
- Standards applicable to the data and data services
- Guidance on the use of the data

[111] https://data.blog.gov.uk/2015/09/25/our-plan-for-data-2/
[112] https://data.blog.gov.uk/2015/09/24/work-of-prog/
[113] https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information
[114] https://data.blog.gov.uk/2015/03/24/progress-on-the-national-information-infrastructure-project/

- Metadata. [115]

There are also many initiatives underway to increase the quality, standardisation, reliability and connectedness of open data. [116] Significant standardisation work is underway to create registers, reliable lists that provide a central stable, reliable and current source of data for specific domains. This provides a stable data service for government and third party providers. [117]

Efforts are also made to harmonise data approaches at the national, county and local levels. Toolkit advice developed at the national level is adapted to jurisdictional variations at the county and local level which enhances its practicality and useability.

A dual focus on public and public sector user needs

Data.gov.uk offers many features for the public to engage with the site, its content and data producers. However government agencies are also recognised as key users of data.gov.uk and site redevelopment work has been focussed on making the data publishing process as simple and efficient as possible. [118] Extensive user consultation and development work has been done to:

- Reduce the complexity of the publishing and updating process
- Outline all requirements in plain English with clear examples
- Improve the location fields and time periods available as descriptive for datasets.

The UK government is also supporting data science literacy for civil servants, recognising the improved service delivery and policy outcomes that will come from greater data use in government. It is also recognised that greater government engagement in data science will also drive improved open data outcomes, by increasing data awareness and use across government. [119]

Comparison with Data NSW

While NSW has the Open Data Policy and a clear open data release mandate from the Government Information (Public Access) Act (GIPA Act) and the Open Data Advocate, the NSW government sector does not have the same open data drivers and enablers as the United Kingdom.

The Information team drives the data program for NSW and has achieved a number of major outcomes through implementing requirements of the NSW Open Data Policy and Open Data Action Plan.

A coordinated approach is being built in NSW to work towards a standardised information infrastructure using enabling legislation like the GIPA Act and the State Records Act,

[115] https://data.blog.gov.uk/2015/03/24/progress-on-the-national-information-infrastructure-project/
[116] These initiatives are outlined on the Data in Government blog at https://data.blog.gov.uk
[117] See https://www.gov.uk/government/publications/registers/registers
[118] See overview at https://data.blog.gov.uk/2017/02/01/making-it-simpler-to-publish-data/
[119] As outlined in https://data.blog.gov.uk/2016/09/02/ideas-to-help-civil-servants-to-understand-the-opportunities-of-data/

collaborative standards development in the spatial data environment and the significant work accomplished under the Information Management Framework. Further work can be coordinated based on the UK approaches to enhance standardisation, and to build further consistency between national and local government approaches.

The Data Analytics Centre is conducting data literacy education programs, available to staff working across the sector. Recent Data NSW survey work is identifying public sector needs and requirements for Data NSW and programs of work are being built to help increase data awareness and use across NSW government.

Key learnings

Key learnings are:

- Significant progress can come from strong governmental drivers and enablers for open data
- Common data policies and governance can built a strong, coordinated and enabling data environment
- A strategic and coordinated approach to data and data quality helps to produce high value datasets
- User design and user experience is important for public and private sector users.

d) Data SF, San Francisco data portal

Background

DataSF is the open data portal for the City and County of San Francisco. It was established in 2009 after Mayor Newsom published an Executive Directive requiring all San Francisco Departments to share their non-confidential datasets on DataSF. [120]

DataSF has evolved to be more than just a repository for government transparency, its goal is to be a data source that builds businesses and improves citizens live through insights. [121] Its mission statement is:

> DataSF's mission is to empower use of data. We seek to transform the way the City works through the use of data. We believe use of data and evidence can improve our operations and the services we provide. This ultimately leads to increased quality of life and work for San Francisco residents, employers, employees and visitors. [122]

[120] https://localwiki.org/sf/Open_Data
[121] http://www.govtech.com/data/DataSF-Relaunches-After-Open-Data-Strategy-Update.html
[122] https://datasf.org/

The DataSF website is organised under the following top-level sections:

Section	Description
Open Data	Provides access to open data datasets through the data catalogue
Showcase	An open data application showcase for demonstrating innovative uses of public data
Publishing	Online submission form (and guidelines) for departments to publish their data
Academy	Links for government staff to training courses in data use, management and processes, including advice on inventory management, setting up dashboards, managing data requests
Resources	Provides open data resources (toolkits, standards, guidelines etc) produced by DataSF
Blog	Called All Things Data, this publicises initiatives, developments and updates on open data

DataSF has a streamlined approach to user search. A 2014 user consultation process led to a site redesign that:

- Simplified the category structure from 27 to 10 categories to improve data discovery
- Made it easier to find data by department
- Visually integrated the various pages to build a consistent user experience across the site
- Simplified, more intuitive and visual help
- Provided quick tips on accessing the API and getting started
- Added an about page with links to our strategic plan and reports
- Enabled easier distinction between original (or source) dataset and derived views, charts, and maps
- Improved metadata
- Built simpler chart creation [123]

[123] https://datasf.org/blog/the-new-datasf/

DataSF has enhanced functionality for dataset viewing and use. This includes a user-interface to every dataset which allows users to easily:

- customise their view of the data by hiding and changing the order of columns
- apply conditional formatting rules, sorting and grouping
- visualise the data with column and bar graphs, pie and donut diagrams, line and area charts, tree maps etc
- join a conversation about the dataset.

While Data NSW provides APIs for some datasets, DataSF boasts API access for all published datasets. DataSF includes a landing page specifically for Developer Resources, and links to a well-designed Socrata development site.

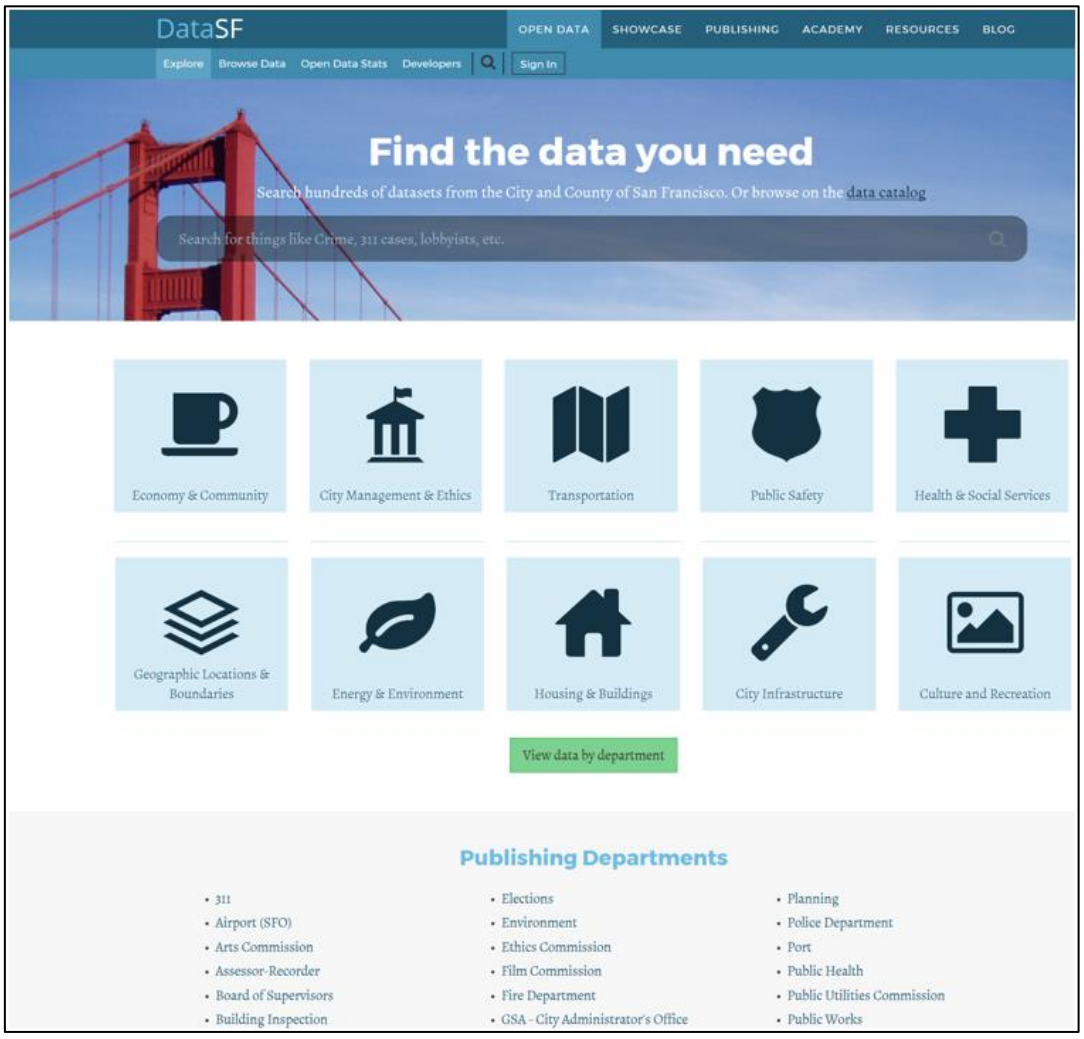
Comparative overview

User experience

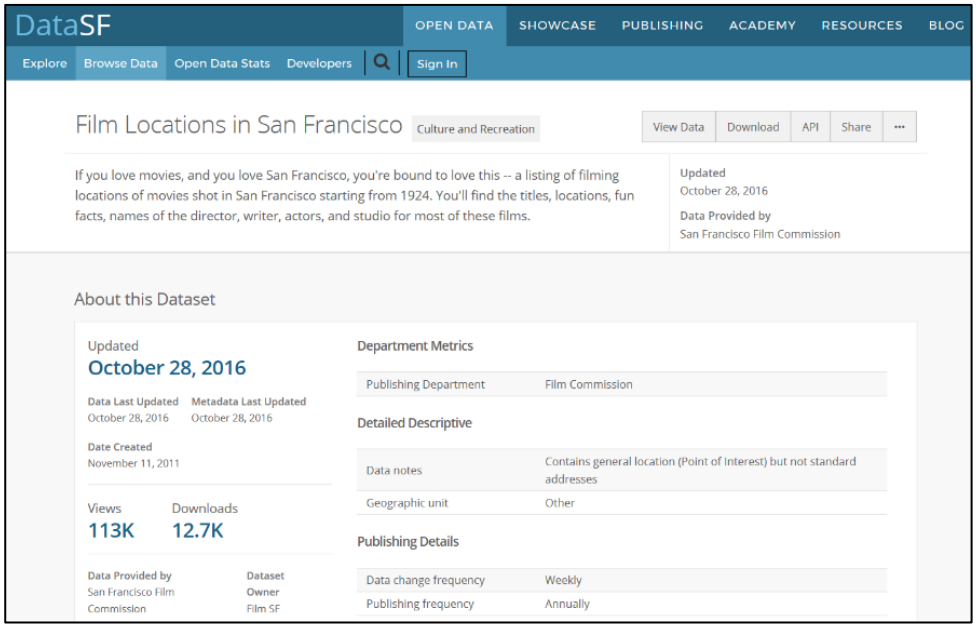
The DataSF portal provides a very well designed user experience for both public and public sector users of the site.

Key characteristics include:

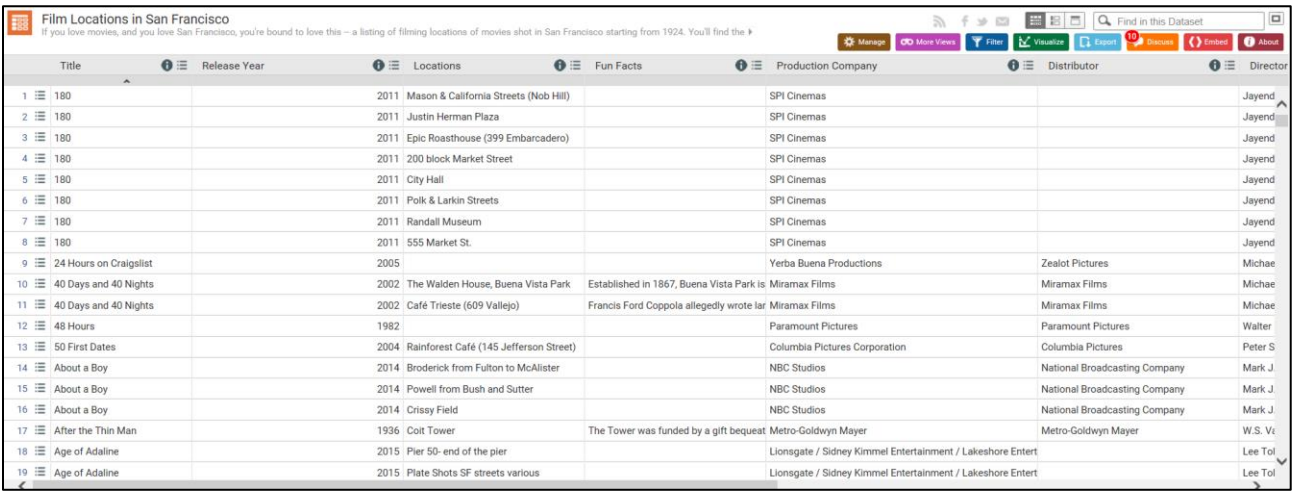
Simple home page layout: Users can easily navigate the website and gain access to the Open Data catalogue



Simple user interface: Users can view data, and find out key information at a glance.

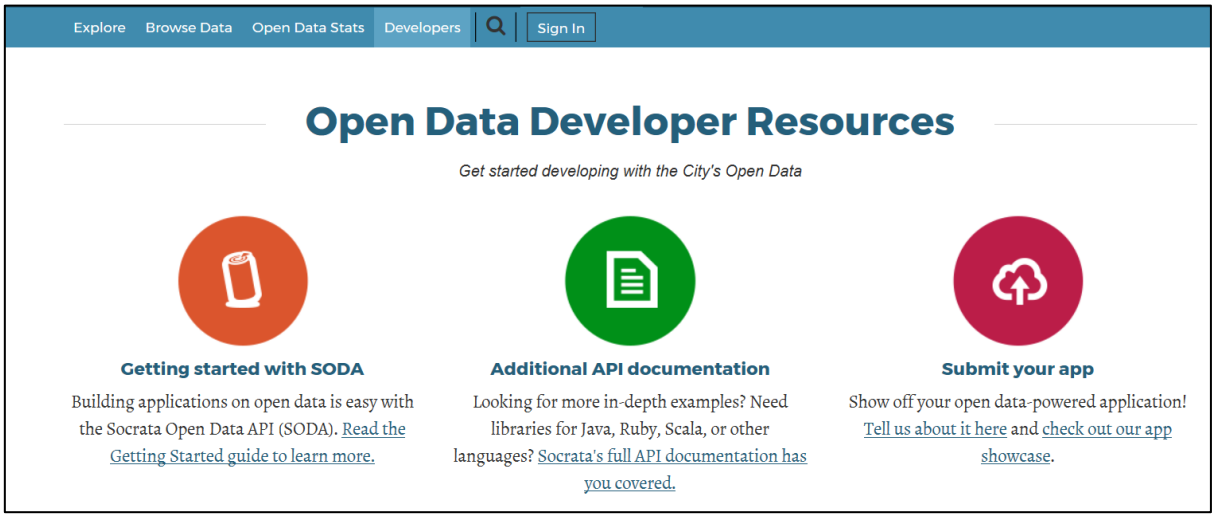


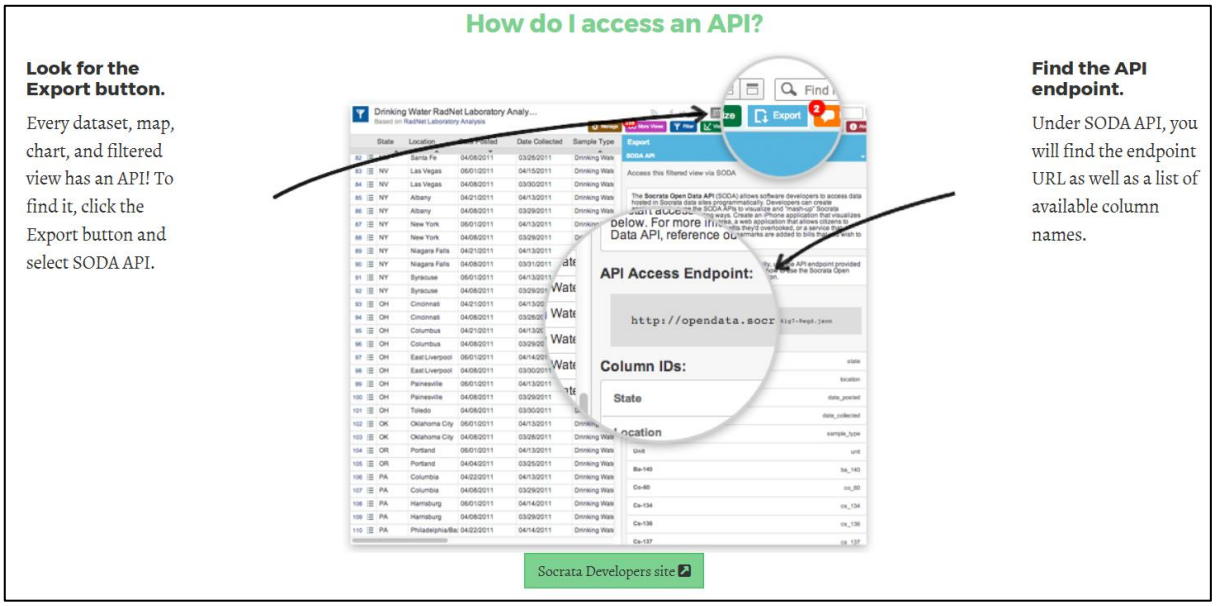
sualise data.



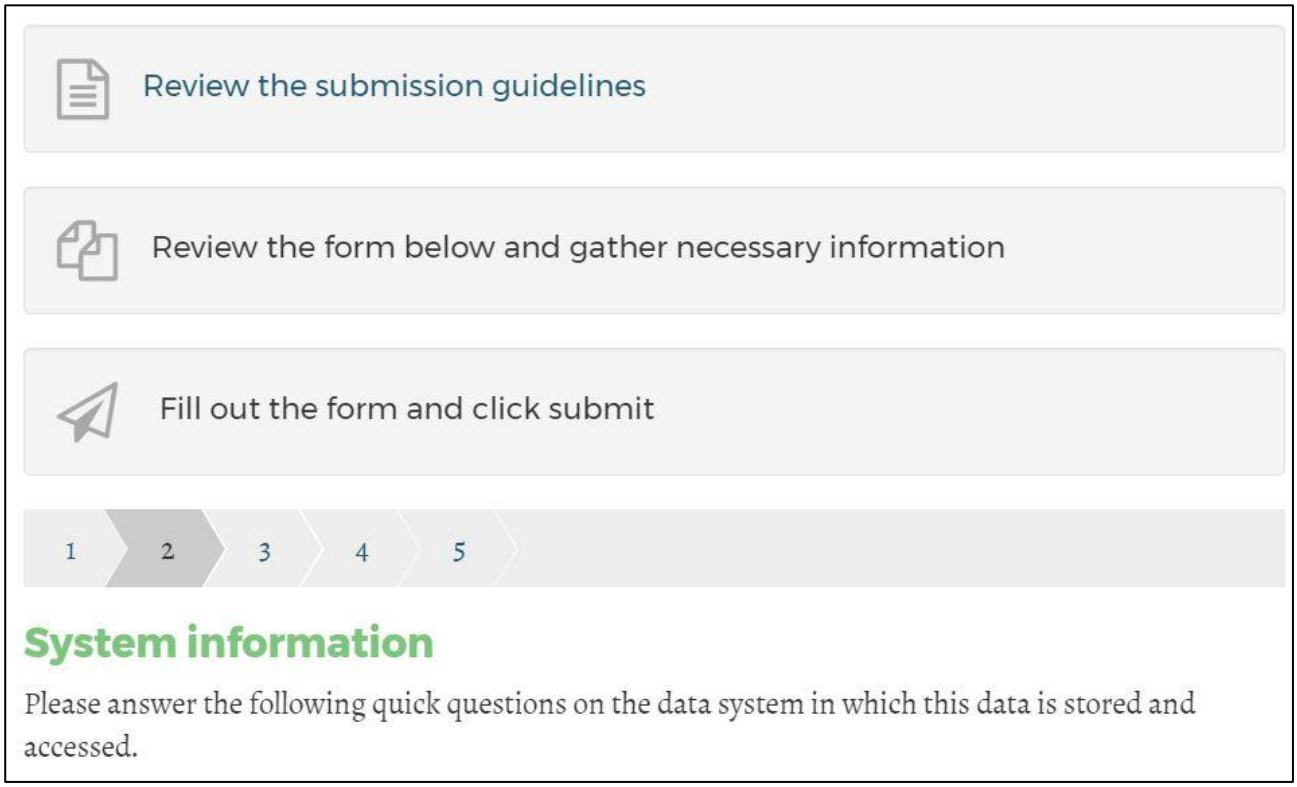
	Title	Release Year	Locations	Fun Facts	Production Company	Distributor	Director
1	180	2011	Mason & California Streets (Nob Hill)		SPI Cinemas		Jayend
2	180	2011	Justin Herman Plaza		SPI Cinemas		Jayend
3	180	2011	Epic Roasthouse (399 Embarcadero)		SPI Cinemas		Jayend
4	180	2011	200 block Market Street		SPI Cinemas		Jayend
5	180	2011	City Hall		SPI Cinemas		Jayend
6	180	2011	Polk & Larkin Streets		SPI Cinemas		Jayend
7	180	2011	Randall Museum		SPI Cinemas		Jayend
8	180	2011	555 Market St.		SPI Cinemas		Jayend
9	24 Hours on Craigslist	2005			Yerba Buena Productions	Zealot Pictures	Michae
10	40 Days and 40 Nights	2002	The Walden House, Buena Vista Park	Established in 1867, Buena Vista Park is	Miramax Films	Miramax Films	Michae
11	40 Days and 40 Nights	2002	Café Trieste (609 Vallejo)	Francis Ford Coppola allegedly wrote lar	Miramax Films	Miramax Films	Michae
12	48 Hours	1982			Paramount Pictures	Paramount Pictures	Walter
13	50 First Dates	2004	Rainforest Café (145 Jefferson Street)		Columbia Pictures Corporation	Columbia Pictures	Peter S
14	About a Boy	2014	Broderick from Fulton to McAlister		NBC Studios	National Broadcasting Company	Mark J
15	About a Boy	2014	Powell from Bush and Sutter		NBC Studios	National Broadcasting Company	Mark J
16	About a Boy	2014	Crissy Field		NBC Studios	National Broadcasting Company	Mark J
17	After the Thin Man	1936	Coit Tower	The Tower was funded by a gift bequeat	Metro-Goldwyn Mayer	Metro-Goldwyn Mayer	W.S. Va
18	Age of Adaline	2015	Pier 50- end of the pier		Lionsgate / Sidney Kimmel Entertainment / Lakeshore Entert		Lee Tol
19	Age of Adaline	2015	Plate Shots SF streets various		Lionsgate / Sidney Kimmel Entertainment / Lakeshore Entert		Lee Tol

Open Data Developer Resources: This page provides open API access to every published dataset, and a landing page for developers with links to API resources:





Simple data upload process: This page provides a simple step by step process to upload data



Data Academy: This page provides training for staff to enhance skills and knowledge of data use.

a range of data related resources.

Stakeholder engagement

Good stakeholder engagement occurs through the DataSF portal. For example quarterly updates are provided to stakeholders on progress against the Open Data Policy. [124]

Government and user stakeholders are kept informed of data publication timetables through the Departmental Publication Plans which can be viewed by delivery timeframes and due dates. These publication plans are also supported by a dashboard indicating whole of sector progress with open data delivery. A full data inventory also provides a list of government datasets. [125]

The DataSF Speaks blog is also an active publishing mechanism to engage with stakeholders and users of DataSF. [126]

Agile approach to data-related work

The DataSF site outlines the agile, iterative and collaborative approach that the Data Team uses in its work. [127]

Principle…	*Meaning…*
Say no to perfection and fail early and often	Do not make perfect the enemy of good. Use small experiments and learn from failures.
Plan for the future and seek institutional homes	Solve problems in a way that creates infrastructure for the future. Find a full-time home for what works.
Use long division	If a problem seems too big, break it into manageable bits. There's always a way to move something forward.
No ugly, old IT	Use modern and lightweight tools. Make designs beautiful and inviting but also a little fun.
Use storytelling and data	Find the people in the data and tell their story. Data without stories is just academic.
Put people, design and needs first (data and technology second)	Start with empathy and use design to guide our work.

[124] See for example the February 2017 quarterly update at https://drive.google.com/file/d/0B-65Qm9J0m0WNDZtdDBfUy1LVVE/view
[125] https://datasf.org/publishing/plans/
[126] https://datasf.org/blog/
[127] https://datasf.org/about/

Continuously learn and listen with humility	Learn from ourselves and others and build on existing work. “Not invented here” attitudes prohibited.
Start with problems, move to opportunities	Start with people's needs and problems and use the chance to demonstrate cool, new stuff for the future.
Be friendly and have fun	Sometimes our work is tough. Take time to breathe, celebrate progress and recognize that we get to work on amazing problems.
If we don’t start now, we’ll never get there	Don’t look back in five years and think “if we had just…”. Every shady street started with a row of saplings.

Key learnings

Key learnings are:

- provide a well-defined user experience for the public and public sector users
- engage stakeholders through multiple channels and make work transparent wherever possible
- take an agile, collaborative and iterative approach to data-related work.

e) Bloomberg Philanthropies ‘What Works Cities Certification’

Background

Launched by Bloomberg Philanthropies in April 2015, What Works Cities is a national initiative to help 100 mid-sized American cities enhance their use of data and evidence to improve services, inform local decision-making and engage residents.[128]

The *What Works Cities Certification* aims to recognise and celebrate local governments that are leading the United States in the use of data and evidence to increase government effectiveness and improve services for residents. [129]

The What Works Cities (WWC) website contains case studies that have achieved the certification. [130] Each of these cities have adopted the WWC Standard. This standard has four components – Commit, Measure, Take Stock and Act – which define how local governments can create a strong foundation for the effective use of data and evidence. [131]

[128] https://whatworkscities.bloomberg.org/about/
[129] https://whatworkscities.bloomberg.org/certification/
[130] https://whatworkscities.bloomberg.org/cities
[131] https://www.bbhub.io/dotorg/sites/8/2017/03/WWC-Standard-Certification-Criteria.pdf

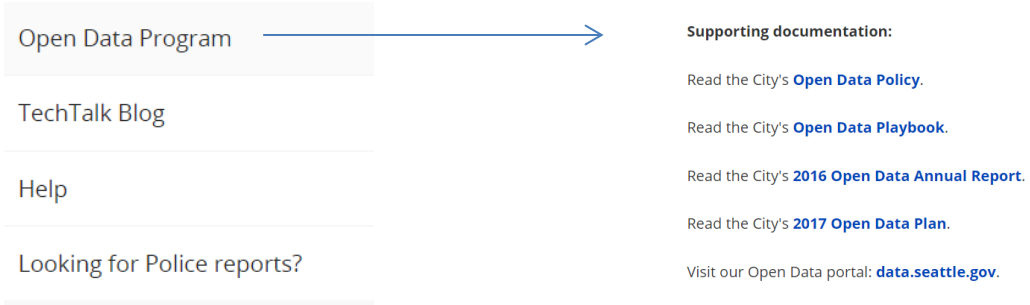
The following examples are drawn from the open data portals from four US cities that have achieved WWC certification: Seattle, Kansas City, Boston, Washington DC.

Please see below for a brief summary of the cities involved in the report and their open data portals:

Seattle, Washington

The City of Seattle was one of the first U.S. cities to pursue open data, creating the first iteration of its open data portal in 2010, data.seattle.gov. Seattle focused on the importance of an open data policy, particularly in the realm of privacy.

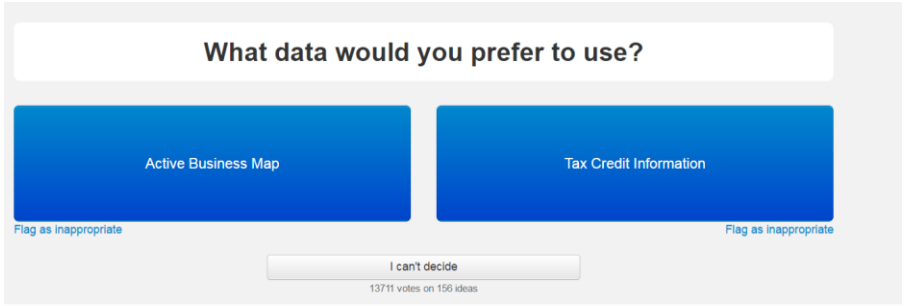
The Open Data Program section of the Seattle Open Data Portal contains a useful collection of supporting documentation to help drive improved open data practice. [132]



Kansas City, Missouri

Open Data KA [133] provides dashboard, map, chat and a comprehensive data inventory to enable citizens to monitor service delivery and engage with government decision-making.

The site is also active in seeking community engagement. [134] It surveys users to gather their requirements for open data. The community is asked to vote on open data ideas submitted by members of the public:

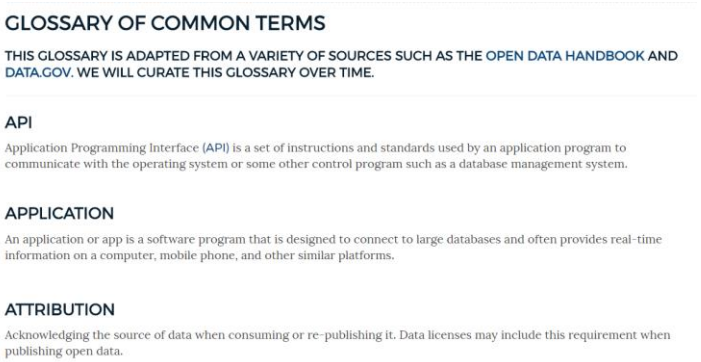


[132] http://www.seattle.gov/tech/initiatives/open-data
[133] data.kcmo.org
[134] http://www.allourideas.org/kcmo/

Boston, Massachusetts

Analyse Boston provides a robust open data, performance, and analytics platform. [135]

Its engaging interface provides a landing page for the portal and displays datasets by topics, popularity, new datasets and modified datasets. It has a strong user experience focus with resources like 'Tips for Users' and 'Showcases'. It also contains a Glossary of Common Terms to make common open data language more accessible to the public.



GLOSSARY OF COMMON TERMS

THIS GLOSSARY IS ADAPTED FROM A VARIETY OF SOURCES SUCH AS THE OPEN DATA HANDBOOK AND DATA.GOV. WE WILL CURATE THIS GLOSSARY OVER TIME.

API

Application Programming Interface (API) is a set of instructions and standards used by an application program to communicate with the operating system or some other control program such as a database management system.

APPLICATION

An application or app is a software program that is designed to connect to large databases and often provides real-time information on a computer, mobile phone, and other similar platforms.

ATTRIBUTION

Acknowledging the source of data when consuming or re-publishing it. Data licenses may include this requirement when publishing open data.

Washington DC

Open Data DC states that 'data's greatest value comes from having it freely shared among agencies, federal and regional governments and with the public to the extent possible when considering safety, privacy and security'. [136]

It provides a range of useful resources in its Tools for Learning and Building area. Each resource provides very accessible, practical and useful advice for data custodians and producers with varying ranges of expertise.

Key learnings:

- Regularly engage with the public through informal surveys on what data is of use and interest to them
- Develop a 'tools for learning and building' area

[135] https://data.boston.gov/
[136] opendata.dc.gov

- Create ‘tips for users’ and an open data glossary to assist first time data users, or to better enable expansion of open data skills.

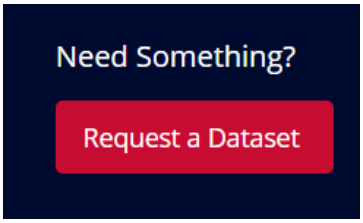
4.2 How well does Data NSW align with government and agency priorities?

4.2.1 NSW government priorities

a) Assessment against requirements of Open Data Policy

The following identifies how Data NSW maps to the principles outlined in the NSW Open Data Policy (2016).

NSW Open Data Policy	
Policy principle	**Assessment of Data NSW**
Open by default and protected when required	*Open by default* Virtually data on Data NSW currently has a default Creative Commons Attribution license (CC-BY) which allows re-distribution and re-use of a licensed work on the condition that the creator is appropriately credited. Only 19 datasets have a ‘licence specified by agency’ licence applied and four have a ‘share alike’ registration which limits commercial re-use potential. *Protected when required* A proportion of datasets on Data NSW have data quality statements. These statements allow agencies to identify when protection, de-identification or de-sensitisation techniques have been applied to protect data. The Data NSW portal does not provide linked advice for agencies on how to protect data when required.
Prioritised, discoverable and useable	*Prioritised* The Policy requires that high-value datasets will be prioritised for release, in line with demand from the public and industry. Data NSW contains a ‘request a dataset’ button to enable the public and industry to request a dataset:



The Open Data Action Plan requires that of high value data including fiscal, extraction industry, environmental and pollution, aggregated and statistical data is prioritised for release.

Specific datasets and portals have been loaded onto or integrated with Data NSW in order to make this prioritised data available including:

- DFSI Budget Data and Mapping the Budget
- SEED Environmental Data Portal
- Open OEH Data Portal
- Workforce Profile and GIPA Act statistics.

Discoverable

To identify datasets it is possible to search for:

- datasets via keyword search
- creating organisation/agency
- groups, such as ‘infrastructure’ or ‘law and order’
- tags applied by custodians such as ‘education’
- data type, such as dataset, image or text
- format, such as XLS, XML, PDF
- licence type.

In addition, to enhance discoverability Data NSW is federated with the Commonwealth, South Australian, Queensland and Victorian data portals, and many NSW-agency based data portals.

Used

Metadata that allows datasets to be found, understood, controlled and managed is published with all datasets.

	Metadata displayed on Data NSW includes Title, Type, Language, Licence, Landing page, Date published, Data Portal, Publisher/Agency.
Primary and timely	*Primary and timely* This requires data to be release as collected at the source, with a high level of granularity and not in aggregate or modified form. The 486 datasets currently available directly on Data NSW are primarily: • PDF-based – 166 • Website links – 74 • XLS – 70 • HTML – 61 • XLSX – 47 • CSV – 28 • XML – 41 • Webservice – 22 • API – 20 • RSS – 15 • Zip – 8 • JSON – 2 PDF-based data is not primary because it has been collated and summarised before being presented in PDF form. For comparative purposes, four per cent of datasets listed on data.gov.uk are in PDF form and not machine processable, whereas 34% of datasets available via Data NSW are in PDF.[137]

[137] The search pages on Data UK and Data NSW list total datasets available and the format view lists total datasets available in PDF form. See https://data.gov.uk/data/search and https://data.nsw.gov.au/data/dataset

Well managed, trusted and authoritative	This requires that users be alerted to the quality and limitations of the data to ensure confidence that it is trustworthy and authoritative. The 486 datasets currently available directly on Data NSW only 25 have data quality statements that would allow users to be alerted to data quality and any limitations.
Free when appropriate	All data on Data NSW is currently free to access, use and reuse
Subject to public input	Currently there is no capacity for users to provide comments or feedback on datasets. There however is the capacity to email data@nsw.gov.au or a range of generic and also name identified dataset contacts associated with specific datasets.

Learnings

- All datasets on Data NSW should have a CC-BY licence
- All datasets should have data quality statements to enhance discoverability and usability and to identify when data has been protected
- Data NSW should link to advice on data protection
- More data on Data NSW needs to be primary and timely. Too many datasets on Data NSW are in PDF form and too few are API-based.
- Data quality tools should be enabled in Data NSW to help agencies to apply data quality tools to datasets that alert users to the quality and limitations of datasets
- There should be capacity for user feedback and direct data@agencyname email addresses for the public to have reliable and ongoing feedback mechanisms, and for agencies to receive ongoing input and feedback about published datasets.

b) Assessment against the requirements of the NSW Government Digital Strategy

On May 23 2017 NSW Government launched its new Digital Strategy.

There are four components to the digital standard for data:

1. Data-informed decision making: All investment decisions, policies and service delivery models are informed by data insights.

2. Open: Data is open, collected and shared in a digital format wherever possible. Agencies open their data in a format that can be readily used.

3. Real time and spatial: Data is published in real time where possible. Visualisation and mapping tools ensure accessible and actionable insights from data.

4. Data sharing: Legislative provisions for data sharing will be observed, as will protections around personal and health information.

In order to achieve the above goals Data NSW can experiment, start and continue to do the following:

Experiment

- Use predictive self-learning tools to measure data quality and benchmarking data quality across the sector.
- Strengthen important protections relating to privacy and security of customer data.

Start

- Enhance and clarify data quality standards.
- Establish the NSW Data Ecosystem to optimise lawful data sharing and use, visualisation and mapping tools across all clusters.
- Enhance quality of data assets such as the single cadastre and address data.

Continue

- Build tools and services for sharing and analysis of data by citizens, industry and government.
- Open up government data while continuing to protect personal and health information.
- Develop robust data governance arrangements.

Learnings

- Data quality and data governance are key enablers of effective digital government
- Open data initiatives are enabled by strong protections relating to privacy and security of customer data
- Use technology to enhance the use and potential of data

4.2.2 Federal priorities

a) Comparison to recommendations from Productivity Commission's Inquiry into Data Availability and Use

In May 2017, the final report of the Productivity Commission's Inquiry into *Data Availability and Use* was published. [138]

In its Inquiry the Commission was required to:

- look at the benefits and costs of making public and private datasets more available
- examine options for collection, sharing and release of data
- identify ways consumers can use and benefit from access to data, particularly data about themselves
- consider how to preserve individual privacy and control over data use.

The Commission's report is referenced here because it contains a number of recommendations that apply to all states and territories. It also outlines extensive recommendations for federal data management practices that will impact on legislative frameworks, data sharing arrangements and data interoperability across all Australian jurisdictions, including NSW.

The following learnings provide requirements relevant to the appropriateness of Data NSW.

Rec No	Commentary
6.1	This recommends that all Australian governments should direct the early release of all non-sensitive publicly funded datasets – both those held by agencies and those held by other bodies that are receiving government funds for their data collection activities. It also states that agencies should report annually on the proportions of their datasets made publicly available, shared and not available for release. This recommendation is aligned to the NSW Open Data Policy but it has significant implications for Data NSW. It will require high level mandates for open data release, collaboration with information regulation bodies on risk assessment and release criteria, and significant data governance and maturity support to enable publication of accessible data to Data NSW. It will also require inventorying of all NSW government datasets, and their possible registration on Data NSW, to enable effective annual reporting.

[138] Productivity Commission 2017, Data Availability and Use, Report No. 82, Canberra. http://www.pc.gov.au/inquiries/completed/data-access#report

6.3	The 'open by default' principle of the Open Data Policy requires that agencies incorporate open data into project, program, service, system and policy design. This recommendation states that: > All Australian governments entering into contracts with the private sector that involve the creation of datasets…should assess the strategic significance and public interest value of the data as part of the contracting process. Where data is assessed to be valuable, governments should retain the right to access or purchase that data in machine-readable form and to subsequently apply any analysis and release strategy that is in the public interest. The recommendation states that all government template contracts should be modified to vest data access and purchase rights in government. Implementation of this recommendation would lead to a routine generation of high value datasets that could be made publicly available through Data NSW and help to achieve the requirements of the Open Data Policy.
6.4	This recommendation maintains that all publicly funded entities in Australia should create data registries. Key components of this recommendation include: • Development of data registries should be achieved by March 2018. • Registries should include metadata, and linked datasets that the entities fund or hold. • Where datasets are held or funded but not available for access or release, the register should indicate this and the reasons why this is the case. • The draft Productivity Commission report recommended that central agencies set objectives to ensure data and metadata are catalogued, searchable and in a machine readable format. This recommendation was removed from the Final Report. • The draft Productivity Commission report recommended that exceptions be applied to highly sensitive datasets. This recommendation was removed from the Final Report, meaning that sensitive datasets should also be registered in data registries.

	Implementation of this recommendation will require rapid expansion of Data NSW, supported by an active inventory program across NSW government to identify and register metadata about datasets, including sensitive datasets. It would also mean that current objectives under the Open Data Policy relating to rolling release schedules and other targets for specific dataset publication would be suspended.
6.6	This states that 'the Australian Government should establish an Office of the National Data Custodian (NDC) to take overall responsibility for the implementation of data management policy, in consultation with all levels of Government'. This organisation will have responsibility for: • oversight and ongoing monitoring of and public reporting on Australia's national data system • recommending designation of National Interest Datasets • accrediting and funding Accredited Release Authorities • providing practical guidance material for ARAs and data custodians on matters such as risk management, data curation and metadata, data security, data deidentification and trusted user models • advising on ethics and emerging risks and opportunities in data use. Implementing this recommendation will potentially impact on the scope, content, metadata and governance models in Data NSW, by creating interoperability and alignment requirements or expectations. It will also important to ensure harmonisation and standardisation where possible between state and federal standards, to minimise confusion, simplify procurement processes and to enable interoperability.
6.7	This recommendation states that the National Data Custodian will streamline approval processes for access to data. It is noted that State and Territory governments could opt into these processes to enable nationally consistent data access and sharing approaches, and streamlined shared between jurisdictions. This will have implications for agencies in NSW, and for data sharing advice that is being developed by DFSI and on agency data publication processes on Data NSW.

6.9	This recommendation states that all 'Accredited Release Authorities must publish formal risk management processes to effectively assess and manage the risks associated with sharing and release of data under their control. Risk management processes should be regularly reviewed and revised to account for new and emerging risks'. This will again potentially create interoperability and alignment requirements or expectations for Data NSW and for open and shared data more generally in NSW.
6.11	This recommendation states that 'The Office of the National Data Custodian should be afforded the power to require an audit of a data custodian's de-identification processes and issue assurance of de-identification practices used'. It is unclear whether any aspects of this requirement will apply to NSW, or to data listed on Data NSW, or whether Data NSW will be required to apply or display assurance statements.
6.17	This recommendation outlines data retention requirements, stating that linked datasets should no longer be routinely destroyed but should be maintained for ongoing access. Implementing this recommendation creates a data retention expectation that will have dataset governance implications for Data NSW. It means that data custodians and DFSI will be required to provide explicit retention statements for datasets and metadata, to ensure that timeframes around ongoing use and accessibility of datasets is explicit and understood.
7.1	This recommends the designation of National Interest Datasets. This could impact on Data NSW by imposing requirements to identify, govern, manage or preserve access to designated National Interest Datasets.
8.1, 8.2, 8.3, 8.5, 8.6	These recommend the establishment of a national Data Sharing and Release Act to provide, to greatest extent possible, a nationally effective, single statute for sharing and releasing data across all Australian governments. This proposed legislation could impact on Data NSW by creating drivers for greater open data release. However, as it also mandates the sharing of data between agencies and jurisdictions, this creates issues as Data NSW has no capacity for more governed forms of data sharing.

	This legislation will apply to all states and territories on an opt-in basis, but it will create interoperability and alignment requirements or expectations for NSW and Data NSW. The proposed national legislation will also mandate that the federal Privacy Act will apply to any state-based data sharing arrangements conducted by mandated Accredited Release Authorities (see recommendation 6.8). It is likely that many cross-sector data sharing arrangements will be co-ordinated by ARAs. This means that governance of Data NSW and other data management environments will potentially need to address both state and national privacy requirements.
9.3	This recommendation states that minimally processed public sector datasets should be made freely available or priced at marginal cost of release, and where data has been transformed, the transformed dataset may be priced above the marginal cost of release. Data NSW has no capacity to charge for data access.
9.4	This states that funding should be provided to agencies for the curation and release of high value, public interest datasets. Although this funding is identified as limited and supplemental, it may create a similar requirement in NSW and the loading of these high value datasets could generate additional access and use of Data NSW.
10.3	This recommendation states that government agencies should adopt data management standards that: • are published on agency websites • are adopted in consultation with data users and draw on existing standards where feasible • deal effectively with sector-specific differences in data collection and use • support the sharing of data across Australian governments and agencies • enable all digitally collected data and metadata to be available in commonly used machine-readable formats, including where relevant and authorised, for machine-to-machine interaction. The recommendation states that these standards should be developed in draft form by the end of 2017, and implemented by all agencies by the end of 2020. Agencies that do not comply will be directed to work towards

	compliance and monitored until they effectively deploy appropriate standards. Implementation of this recommendation will impact on Data NSW by creating interoperability and alignment requirements or expectations with national data management frameworks. Improved national agency-based data quality and maturity will also lead to a significant increase in the number and quality of datasets made available on data.gov.au. This could contrast with the fairly static growth rates of data on Data NSW. Data quality is a key requirement of the Digital Strategy, however the implementation timetables for the Digital Strategy do not align with the aggressive timeframes outlined in this recommendation.

Key learnings based on Productivity Commission report:

- enabling data sharing, as well as enabling access to open data is a key recommendation
- the directive to release all available datasets will require high level mandates for open data release, collaboration on risk assessment and release criteria, data governance and maturity support to enable publication of accessible data to Data NSW, and inventorying of all NSW government datasets, and their possible registration on Data NSW, to enable effective annual reporting
- open data requirements should be built into project, program, service, system and policy design wherever possible
- data management standards are key enablers of effective data frameworks
- comprehensive data registries are key enablers of open government and innovation
- Commonwealth government will produce a range of policies, guidance, processes relating to data, data management, data risk management, data sharing and metadata

4.2.3 Assessment against emerging trends

a) The Internet of Things (IoT) - *Enabling the Internet of Things for Australia* report

The Internet of Things (IoT) utilises sensors and actuators connected by networks to computing systems to transfer data and gain real time insights. It is expected that rapidly

growing demands for IoT will lead to significant innovations in telecommunications and ICT infrastructure markets and predicted that IoT will drive the smarter use of infrastructure, leading to efficiency gains and business growth. The McKinsey Global Institute's 2015 report *The Internet of Things: mapping the value beyond the hype* concluded that globally, IoT could generate up to USD 11.1 trillion per year in economic value by 2025. [139]

Open data has been identified as a key resource for realising the potential of the Internet of Things. The report *Enabling the Internet of Things for Australia* states that 'access to open government data is a major enabler for IoT and for industry'. [140] The report identifies that the open data market in Australia is critical because it is opened, federated, trusted, analysed, secured, standardised, valued, wholesaled, shared and guaranteed. These characteristics are critical for realising the full potential of IoT and providing the stability for economic and industry investment in IoT-based businesses. The report recommends that Australia takes the lead in opening datasets [141] and in enabling greater data sharing to underpin IoT acceleration. [142] The report specifically states that:

> 'By reinvigorating open data policies, there would be a contribution to Australia's cumulative GDP growth of $16 billion per annum or around 1% of GDP over the next five years. Australia could take the lead in opening up data sets and enabling sharing which can underpin IoT acceleration.' [143]

Data silos are identified inhibitors of IoT acceleration [144], whereas open data policy and data sharing principles are identified enablers of IoT.

The report identifies that frameworks for sharing data are also important for opening usage and value in IoT, and governments can be key players in defining these frameworks. Gaps in regulations governing data ownership and privacy are identified as issues affecting e-health in Australia, that could affect other IoT sectors. [145] Resolving these data concerns will require coordinated government action. The report also identifies that 'Frameworks for sharing data are proving to be useful in opening usage and value. This is occurring in some sectors and in are in discussion and development at Government level in many countries.' [146] Further development of data sharing frameworks across government is therefore a necessary enabler of IoT.

Capture, storage, management, security and analytics of the vast amounts of data generated by IoT devices is also a critical issue. The capture and storage of IoT data is identified as having 'profound implications and opportunities' [147] that could impact on open data platforms like Data NSW. The use of this real-time data could have profound benefits,

[139] http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world
[140] Heydon, G. Zeichner, F. Enabling the Internet of Things for Australia. (2015) Communications Alliance Limited. Retrieved from http://www.commsalliance.com.au/__data/assets/pdf_file/0007/51991/Enabling-the-Internet-of-Things-for-Australia.pdf, p77
[141] See Observation 33 of Enabling the Internet of Things for Australia, p77
[142] See Observation 21 of Enbling the Internet of Things for Australia, p66
[143] See Observation 33 of Enabling the Internet of Things for Australia, p77
[144] See Observation 19 of Enabling the Internet of Things for Australia, p89
[145] See Observation 26 of Enabling the Internet of Things for Australia, p72
[146] Frameworks for sharing data are proving to be useful in opening usage and value. This is occurring in some sectors and in are in discussion and development at Government level in many countries.
[147] See Observation 15 of Enabling the Internet of Things for Australia, p61

but standards, infrastructures and capacities for this data use are still only nascent and again, are likely to involve some degree of government-based involvement to enable the standardisation and stability necessary to foster business innovation.

Learnings based on *Enabling the Internet of Things for Australia* report

- Recognise the economic potential of platforms like Data NSW that make stable, valuable, standardised and guaranteed sets of government data available for business and community re-use
- Understand the need to enable greater data sharing
- To facilitate better sharing, it is important that data ownership and privacy requirements are clearly defined.

4.2.4 Results from user surveys and interviews

In May 2017, two surveys were conducted by DFSI to identify perspectives on the appropriateness of Data NSW, the extent to which it addresses an identified need and how well it aligns with government and agencies priorities. It also sought to identify perspectives on are users being reached as intended and sought to assess user satisfaction. These efficiency measures will be reported in the following section.

The 30 question *Key Stakeholder Survey* was sent to key data contacts in agencies that are significantly engaged with open data in NSW, either as producers and custodians of open data or as key stakeholders in open data. The survey was sent to nominated data contacts in:

- Office of Environment and Heritage
- Department of Planning and Environment
- Transport for NSW
- State Archives and Records Authority
- Information and Privacy Commissioner
- Department of Primary Industries
- Department of Financial Services, including Spatial Services.

This survey had a small sample size, but targetted an informed and engaged constituency. Six responses were received.

The 14 question *Data NSW User Survey* was distributed to information professionals working in NSW public sector through two existing networks, the Information Management Community of Expertise and the Privacy Professionals Network. The information

professionals targeted by this survey are aware of data and data strategies in their organisations, but are unlikely to be directly responsible for open data strategies.

This survey was also promoted through the Data NSW blog [148] and reached a number of Data NSW private sector users through this channel. A range of responses were received from the public and private sectors. This survey received 37 responses.

The responses from this survey show that, while Data NSW has served some specific needs and helped some agencies to achieve progress with their open data, ongoing work is required to build an engaged and sustainable open data environment in the NSW public sector.

A summary of survey results and comments is provided below. A full list of questions from each survey is at Section 10 Appendices.

a) Key Stakeholder survey – Data NSW implementation appropriateness

Key findings from this survey relating to Data NSW implementation appropriateness, and the extent to which Data NSW meets identified needs or government and agency priorities are as follows.

Understanding of Data NSW

- 60% of respondents reported that there was good awareness in their organisation of Data NSW. 20% reported that there was very good awareness, and 20% reported that there was limited awareness
- 60% reported that staff in their organisation understand the purpose of Data NSW. 20% reported that staff did not understand the purpose, and 20% were unsure. One agency commented: 'Difficult to know if it is a metadata portal or a data portal…also, all very manual if you have to enter in metadata statements to Data NSW when you already have your own portal.'

Use of Data NSW

- 80% of respondents had uploaded data to Data NSW in the last 12 months, and 100% had uploaded in the last 12-24 months
- 50% had used datasets on Data NSW for project, policy or research work or to access real-time data feeds; 33% reported that none of the datasets on Data NSW were relevant to their work
- 80% have used the Information Asset Register to register information
- 60% of the responding agencies have an Open Data Strategy, 40% do not

[148] https://data.nsw.gov.au/blog/have-your-say-data-nsw

Main organisational challenges with open data

When asked, respondents reported that their organisations' main challenges with open data are:

- 'Managing risk of data misuse'
- 'Convincing agencies to release data'
- 'People understanding the information they have is data to be uploaded in raw format'
- 'Finding the time/resource to prepare data for release'
- 'Different contractual agreements around projects'
- 'The resources and time to make data open'
- 'The privacy concerns about making data available'
- 'Internal data/information governance'
- 'Publishing as web services'.

When asked how Data NSW could be improved to address these challenges respondents reported:

- 'More scope for data visualisations alongside raw data'
- 'Get a clear direction/objective of Data NSW – decide if it should be revamped or removed.'
- 'Clearly message why do we need to use Data NSW vs not or other entities'
- 'I'm not sure it can, the responsibility really lies within agencies'
- 'Work with the risk based model to help provide guidelines on how to do it better'
- 'Provide guidelines on the privacy aspects'
- 'Should provide more automation'
- 'Help agencies to mature their information management by providing an IAR that connects to Data NSW'.

When asked about the main challenges that their organisation has with shared data, respondents reported:

- 'Datasets not available in usable formats'
- 'Finding time/resource to invest in preparing data and keeping up to date'

- ‘Different contractual arrangements for projects with joint funding’.

Learnings based on the findings of the Key Stakeholder survey are:

- Key stakeholders are engaged users of Data NSW and have significant insights to share on open data enablers and inhibitors
- It is important to provide clear role and purpose statements for Data NSW. Communicate these in all messaging to ensure its role is clear to all relevant communities.
- Toolkits and advice that address governance, privacy, security and risk are important enablers of open data.

b) Data NSW user survey

Key findings from the *Data NSW user survey* are as follows.

Agency awareness of Data NSW

- Responders to the user survey reported that there is not strong awareness of Data NSW. 43% of respondents said some staff in their organisation are aware of Data NSW. 30% said hardly any staff are aware and 19% said only staff who work with data are aware of Data NSW. 8% reported many staff are aware of Data NSW.
- When asked if people in their organisation understand the role and purpose of Data NSW, 60% said no, 32% said yes and 8% were unsure. Respondent commented:
    - ‘I think there is potentially good understanding by a small number of people and limited understanding and/or awareness amongst the majority.’
    - ‘We are very thankful that we now have excellent access to good administrative data that was often not accessible in the past.’
- 18% of respondents reported that their Executive understand the role and purpose of Data NSW, while only 9% reported that business areas understand the role and purpose of Data NSW.

Agency commitment to open data

- 82% of respondents said their organisation was committed to open data release. 18% said their organisation was not committed to open data release.
- Some of this support was tempered however, and some stated support may reflect organisational commitments but not organisational actions. Respondents comments included:

- ‘We are supportive of open data however we have concerns regarding data accuracy. Also data being used for purposes beyond if its ‘fitness for purpose. Data currency is another issue. It may take months for updates from the data creator to filter down to users.’
- ‘Inappropriate release of data sets has the potential to impact adversely and severely on individuals or population segments. Without stronger governance, assurance and remediation pathways than are currently in place, not all datasets that might usefully be released or shared (either with unlimited sharing or closely governanced sharing structures such as maintained by the ABS) should be shared.’
- ‘Committed at Executive level, perhaps not seeing in action or embedded in the culture of staff generally.’
- ‘Don't understand why Government organisations would want to spend time and energy to hand over data for free to third party developers who then go and make money off it.’

Agency use of Data NSW

- 16 respondents reported uploading data to Data NSW in the last 3-5 years. 17 reported that they have not uploaded data. The key reasons given for this, in rank order are: we don’t have appropriate resourcing (41%), we don’t have any appropriate data (35%), we have privacy and safeguarding concerns (35%), we don’t have appropriate skills (35%), we are currently working to get some datasets ready for publication (18%), uploading data to Data NSW is difficult (6%).
- Of the respondents who had uploaded datasets, only 13% knew how they had been used by industry, government or the public

Learnings relating to appropriateness from Data NSW user survey:

- There is limited awareness of Data NSW among executive and business staff.
- While there is stated support for open data initiatives, this is not leading to open data practice in many organisations.
- Agencies are experiencing a lot of transformational change. Try to build open data initiatives as integrated parts of other tools or drivers, such as the Digital Strategy, to minimise impact and maximise benefit for agencies.
- Providing better metrics and reporting on dataset use can help to report on benefits realisation and build the case for open data.

5. Key findings: Implementation – Efficiency

5.1 Is Data NSW being implemented appropriately?

This section examines national and international best practice research into open data enablers and inhibitors to determine whether Data NSW has been implemented efficiently.

5.1.1 National and international best practice open data enablers

a) Conditions Enabling Open Data and Promoting a Data Sharing Culture: A report for the Information and Privacy Commission of New South Wales

This UNSW Law Report was commissioned by the NSW Information Commissioner and NSW Open Data Advocate to provide contemporary insights to support the promotion of Open Government and open data.[149] The report analysed legislation, policy, regulatory settings, roles and responsibilities for leadership, culture and operations relating to open data in in leading jurisdictions as identified in the *Open Data Barometer Report* – United Kingdom, United States, France, Canada and New Zealand.

The report identifies that successfully implementing an open data agenda requires a suite of five mutually-supporting actions. These are:

- Leadership at the national and sub-national levels
- Adopting appropriate legislative, policy and regulatory settings
- Cultural change in the public sector and broader community
- Collaboration between government agencies and external stakeholders
- Communication of early successes.

In looking at success factors from open data initiatives across the world, the report identifies the following as key enablers of an open data culture:

[149] Dr Alana Maurushat, 2017. UNSW Law. Conditions Enabling Open Data and Promoting a Data Sharing Culture: A report for the Information and Privacy Commission of New South Wales; http://www.ipc.nsw.gov.au/sites/default/files/file_manager/Conditions_Enabling_Open_Data_Report_Final.pdf

Build a shared understanding that open data is:	• machine-readable data in a standardised format delivered with a standardised licence • supported by strong data governance frameworks that have clearly articulated roles and responsibilities, and provide detailed information and guidance around processes and tools • enabled by privacy frameworks and data assurances
Build a national data portal that:	• use standardised licenses • use standardised metadata • inventory and catalogue datasets • make limitations and restrictions on datasets clear • suggest optimal uses and tools for maximising the use of datasets
Build culture and collaboration	• throughout the data lifecycle • with open data champions • through skills exchanges and mentoring across government sectors
Build data stories	• start with pilot studies or specific problems • incentivise other agencies to open data through compelling success stories or use cases • realise that data on its own will not lead to improved data sharing and an open data culture

The report outlines a range of specific learnings for NSW to implement in order to enable an open data culture. These learnings are grouped in the following categories:

- leadership

- legislative, policy and regulatory
- culture and collaboration
- operational.

Leadership domain

Required action…	*In order to…*
Ministers to issue open data directives	Mandate that their departments release open data in machine readable formats
NSW government adopt the Open Data Charter	To publicly commit to open data agendas and targets
Qualitative and quantitative impact measurements	Create international tools and benchmarks to assess the impact of open data initiatives

Legislative, policy and regulatory domain

Required action…	*In order to…*
Identify core datasets	Understand which datasets are important economic drivers for NSW's future
Mandate open release of successful GIPA requests	Streamline the routine release of accessible information in machine readable formats
Fund the opening of high value datasets	Ensure agencies are supported to release publicly valuable information in machine-readable formats
Mandate departments to open specific datasets	Provide continued growth of accessible and valuable datasets
Mandate departments to create machine-readable standardised formats for datasets	Allow for analytics and linked data applications
Mandate metadata standards	Standardise all datasets, licenses and machine-readable formats, to ensure consistency and enable innovation
Publish a complete catalogue of all datasets	Including datasets that are restricted, to enable identification and access of all government data

Develop detailed guidelines on anonymisation and de-identification	To develop a balanced approach to the risk of harm resulting from any re-identification
Develop model Privacy Impact Assessments	To provide practical guidance on balancing risks and benefits of data release
Develop a standardised 'Open Government License'	To promote consistency and ensure compatibility with the Creative Commons License

Culture and collaboration domain

Required action…	*In order to…*
Influence agency and staff attitudes	Build a supportive open data culture across government
Elevate community understanding of open data	Engage the community in open data availability and use
Cultivate data sharing between jurisdictions	Enable service innovation, improved policy development and national consistency

Operational domain

Required action…	*In order to…*
Develop greater public sector capability in open data	Drive improvements in open data use and availability
Improve the management of open and shared data	Understand the enablers within government and between government and the private sector
Share stories of successful programs and outcomes	Help build a self-sustaining open data culture
Develop problem statements that can assisted with open data	Develop compelling use cases for open data
Commence a pilot study to solve a problem or tell a story	Develop compelling use cases for open data

Key learnings from the *Conditions enabling open data and promoting data sharing culture* for the appropriate implementation of Data NSW:

- culture and collaboration are the most important factors for developing a sustainable open data culture
- open data has to be supported by strong governance frameworks
- data must be used to solve problems and sharing these stories will help to build a strong open data culture.

b) Open Data Barometer Report 2017

The Open Data Barometer Report is a big-picture study that take a systematic approach and makes quantitative comparisons. Its rigorous approach allows for identification of trends and patterns and its research produces hard evidence to support policy. It is an international measure of open data that assesses readiness, implementation and impact of national open data initiatives. The report has a well-designed methodology and is a concerted effort of 150 researchers and government representatives from across the world. [150]

The Open Data Barometer Report 2017 noted that 'Government data is typically incomplete and of low quality'. [151] Open data portals were identified as playing a role in this, in that they require agencies to perform additional manual processes in order to release open data. The report found that:

> Government data is usually incomplete, out of date, of low quality, and fragmented. In most cases, open data catalogues or portals are manually fed as the result of informal data management approaches. Procedures, timelines, and responsibilities are frequently unclear among government institutions tasked with this work. This makes the overall open data management and publication approach weak and prone to multiple errors.

As a consequence the report recommended that open data release be built as a business as usual process in all government agencies. A key recommendation of the report was that 'governments must decentralise open data across all agencies and departments':

> In order to guarantee long term sustainability of open data, all government data management practices and systems must be designed with openness in mind from the very beginning of the data management process. It is imperative that governments do not see opening data as an additional step at the end, but as something to be integrated throughout the whole of government. We recommend that governments review their data governance processes in full and also embed automated data publication processes in their IT systems.

[150] The methodology for the Open Data Barometer is outlined at http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-Methodology.pdf
[151] Open Data Barometer 4th Edition — Global Report, May 2017 | The World Wide Web Foundation http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf pages 14-15

> This will ensure the latest and most complete version of datasets is always available to the public directly from the source and will reduce reliance on manual uploads to one single central catalogue.
>
> As it stands, open data portals should be considered as a temporary workaround in order to enable access to government data while a more consistent solution is implemented. This might be that data is published in an automated way, as outlined above, on the websites of relevant departments, with a further automated step to populate a central portal in real-time. [152]

The following is a list of the issues and learnings that have been outlined in the report, based on the report's world-wide assessment of open data practices.

Issues:

1. This year's report shows that governments are slowing and stalling in their commitment to open data. Nine out of ten government datasets are not open.
2. Government data is usually incomplete, out of date, of low quality, and fragmented. This makes the overall open data management and publication approach weak and prone to multiple errors.
3. Sustained political will is what makes or breaks the success of open data. Political will needs to be translated into strong legal and policy foundations.
4. Governments are not publishing the data needed to restore citizen's trust. Open data portals often do not contain the data people really want and need. Money should be spent on opening up datasets that people need.
5. Few open data initiatives actively promote inclusion and equity. Data policies must also be inclusive by design, in order to harness the potential of open data to improve equality and social outcomes.

Open Data Barometer Report 2017 recommendations

1. Government-held data must be open by default and follow the principles set out in the Open Data Charter. Right to Information (RIT) laws should be revised to provide for proactive disclosure that guarantees non-personal government data will be open by default, available in machine-readable formats, and published under open licenses that allow the data to be re-used.
2. Governments must decentralise open data across all agencies and departments. It is imperative that governments do not see opening data as an additional step at the end, but as something to be integrated throughout the whole of government.

[152] Open Data Barometer 4th Edition — Global Report, May 2017 | The World Wide Web Foundation http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf, p15

Governments should review their data governance processes in full and also embed automated data publication processes in their IT systems.

3. Governments must adopt the Open Data Charter to ensure open data practices are embedded beyond political mandates. Adopting the charter will provide strong policy foundation and a consistent data management strategy and practice.

4. Governments must consult citizens and intermediaries when prioritising which open data to publish first. Governments must work with data intermediaries to find out exactly what data and information citizens need to address their problems and improve public services.

5. Governments must invest in using open data to improve the lives of marginalised groups. Data policies must also be inclusive by design: data collection, data design, data access, data use.

Learnings from the Open Data Barometer report 2017 for the appropriate implementation of Data NSW

- Ensure all open data is machine readable with an open license
- Engage citizens in dataset release prioritisation and design
- Embed open data in system design and contract management

c) Literature review: European Commission report, Recommendations for open data portals: from setup to sustainability

This report was commissioned by the European Commission to identify what is needed to enable the sustainability of data portals. The report states that 'Open data portals are a critical part of our data infrastructure: they connect data holders with data users, who in turn create services that citizens and businesses benefit from and rely on'. [153] It identifies however that most open data portals across Europe do not have coherent strategies to support their sustainability. As a result:

> Portal owners now face a series of interlinked problems: limited funding, challenges instigating re-use of data, data quality issues, securing political support for publishing relevant data, and scaling up with legacy systems and operating structures. This means many portals struggle to remain fit for purpose.

The report assessed eleven European data portals and noted that generally these portals were set up:

[153] Sasse, T. Smith, A. Broad, E. Tennison, J. Walls, P. Atz, U. (Open Data Institute). Recommendations for Open Data Portals: from setup to sustainability. (2017) European Commission. Retrieved from https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf

- as part of a specific Open Data initiative, and administered by the team responsible for the initiative
- without the need for a comprehensive business case or user research to help make the case for funding (because Open Data was a strategic political objective)
- separate to (in terms of governance and staffing) digital functions and strategies for government, although with obvious overlap. [154]

To become sustainable, the report identifies that open data portals and their governance need to be embedded as 'business as usual', directed by a strong business plan and with a key basis in community engagement.

Open data portals also need strong supporting processes including processes to govern:

- operational service management that ensures the portal operates as it should
- on-boarding new users onto the portal, which is particularly important for publishers
- training and knowledge transfer to ensure that experience in the running of the service is passed on.

The report identifies five key areas where portals need to plan and manage sustainability:

Sustainability areas	**Key learnings**
Governance	• Have a business plan and a clear governance structure • Bring publishers and data users together to address specific challenges • Build responsiveness to government priority changes into governance structures • Create hard levers to set and enforce data quality and metadata standards, and pursue data updates from publishers • Create a non-ministerial leadership role to champion data publication and respond to issues
Finance	• Be open about the funding strategy, so that publishers and users can identify future needs, use cases and potential funding shortfalls • Ensure priorities align with those of funding sources

[154] https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf, p17

	• Set funding strategies and budgets • Undertake research on the impact of the portal's current or potential activities, to support a business case for future funding
Architecture	• Select open source solutions and solutions that offer archiving/downloading options for all data published via the portal • Contribute to the development of standard APIs that can be used across all open data platforms for sharing, summarising and presenting data • Build links to data held in other portals, to create more data opportunities for your users • Even if not responsible for the publication and maintenance of data, research user needs and preferred data formats to drive data improvements • Build on recognised standards to foster interoperability and comparability of metadata
Operations	• Manage publication operations to support different types of publishers, from small to large scale, enabling automation where possible • Manage technical operations to include effective monitoring and reporting systems for inaccessible data, preferably through publicly accessible lists for users to track progress • On-board new end users, publishers and monitors with effective User Experience design, clear publications processes, feedback loops and training • Automate functions to ensure seamless integration of a diversity of data sources, increase user friendliness and limit overheads for stakeholders • Capture and share lessons learned, and be open to best practices and standards developed by other portal operators
Metrics	• Choose metrics that help to benchmark data publisher performance, but do not rely on one metric eg quantity.

	Combine quantity metrics with data quality and engagement measures • Choose metrics that help potential data users find data that is suitable for them to use. Evaluate whether the metrics chosen are meaningful or potentially misleading • In measuring quality of open data, take into account metadata and contextual information to increase user understanding and engagement • Overcome challenges in automating metrics by adopting standard language and terminology, publishing processes and metrics for data quality and reuse

Key learnings from the Recommendations for open data portals: from setup to sustainability report for the appropriateness of Data NSW:

- Develop funding strategies and budgets that support the portal's planned and future needs and directions
- Collaborate with other portal owners to share knowledge and lessons learned
- Build on recognized standards to foster interoperability, ease of tool development and data re-use

d) Evaluating the Quality of Open Data Portals on the National Level

Czech researchers Renata Máchová and Martin Lnenicka have developed a data portal gap analysis framework. [155] This framework was based on a systematic literature review and emphasised quality dimensions and a set of general characteristics of datasets and their metadata quality. Data NSW fairs strongly against the assessment. However, the success of data portals are largely dependent on the content (data maturity, quality and relevancy of the datasets), a factor not addressed in this gap analysis.

GENERAL CHARACTERISTICS OF AN OPEN DATA PORTAL		
List of metrics	**Description of the requirements for the quality evaluation**	**Data NSW compliance**

[155] Renata Máchová, & M. Lnenicka, Evaluating the Quality of Open Data Portals on the National Level, *Journal of theoretical and applied electronic commerce research*, Volume 12 No 1, 2017, accessed via http://dx.doi.org/10.4067/S0718-18762017000100003

1.0 Technical dimension		
1.1 Authority and responsibility	Portal provides information about the authority, which hosts the portal and the governance model or institution framework supporting data provision models.	
1.2 Data Management system	Portal provides information about the data management system used to power the system.	
1.3 Language	Portal provides more than one language version to gain more users (attention) and improve the overall quality of the portal.	
1.4 Free of charge	Portal provides all datasets and services free of charge and without restrictions under open licenses.	
2.0 Availability and access dimension		
2.1 Number of datasets	Portals should provide the number of datasets they provide	
2.2 Number of applications (reuses)	Portals should provide number of applications developed based on the open data re-used	
2.3 Search engine filter	Portals should adopt and make visible an overall organisation structure and provide strong dataset search capabilities and selection tools using different criteria for browsing through categories and browsing through filters.	
2.4 API	API Portals should provide API for stakeholders to develop applications using open data	
2.5 User Account	Portals should support user account creation in order to personalise views and information shown	
2.6 Thematic categories	Same tags should be used to classify data of the same type and category	

2.7 Tags (keywords)	The portal should clearly distinguish categories (themes) from tags (keywords)	
3.0 Communication and participation dimension		
3.1 Forum Feedback	Portals should provide an opportunity to submit feedback on the data from the users to providers and forum to discuss and exchange ideas among the users	
3.2 Request form	Portals should provide a form to request or suggest new type or format type of open data.	
3.3 Help (usability)	Portals should include high quality of documentation and help functionality to learn how to use the portal and improve the usability	
3.4 Frequently Asked Questions	Portals should include FAQ to help resolve any issues	
3.5 Social Media	Portals should be connected to a social media platform to create a social media distribution channel for open data. Users and providers can inform each other about what they did with and re-used from a dataset.	
4. General Characteristics of dataset		
4.1. Title and description	Datasets should be provided together with their description and also how and for what purpose they were collected	
4.2. Publisher	Publisher Datasets should be provided together with their publisher to verify authenticity of their source	
4.3 Release date and up to date	Datasets should be explicitly associated with a specific time or period tag. All information in the dataset should be up to date	

4.4 License	Datasets should provide license information related to the use of the published datasets. Datasets that doesn't explicitly have an open license are open data	
4.5 Geographical coverage	Datasets should be determined if the coverage of data is on the national, regional or local level	
4.6 Dataset URL	Datasets URL should be available for each dataset	
4.7 Dataset (file) size	Size should be available	
4.8 Number of views (visits)	Total number of online views should be available for a dataset	
4.9 Number of downloads	Total number of downloads should be available for a dataset	
4.10 Machine - readable formats	Datasets should be provided in formats that are as convenient, easy to analyse and modifiable as downloadable files in well-known formats	
4.11 Visualisations	Visualisation capabilities should be provided. e.g. as visualizations in charts or visualizations in maps	
4.12 User rating and discussion message	Datasets should provide capabilities allowing to collect user ratings and comments on a dataset or to discuss conclusions based on data use	

Key learnings from the Data portal gap analysis framework for the appropriateness of Data NSW:

Mapping against this framework shows that Data NSW compares very favourably to this comprehensive international benchmarking tool.

e) Apps4NSW survey, October 2015

In October 2015 a survey of 102 NSW-based IT developers was commissioned by the Department of Finance, Services and Innovation. [156] This survey was designed to identify the drivers and concerns that the development community has in relation to working on innovation challenges with Government. However the survey also contains interesting insights into the data-related needs of the development community, and their interest in working with government data.

71% of survey respondents said they were likely or very likely to participate in creative development events using government data.

86% of respondents also reported that they would like greater opportunities to work with other data sources, not just government datasets.

Respondents also reported that they want to access data that is clean and ready to use, accessible in a standard machine processable format or through an API. In addition, whether data is point in time or live data, participants want access to continuous data updates (where applicable) to ensure the currency of their solutions.

To support innovation, developers want government data websites to provide:

- data themed into broad topics
- data linked with additional insights to help explain its usefulness
- searchable data
- real life stories that support the data.

The developers reported that they were excited by open data, but to fully engage with its use, respondents identified that they want the capacity to engage with data that enables them to develop solutions to real community problems, or to build insights into genuine community issues.

Key learnings from the Apps4NSW Survey for the appropriateness of Data NSW:

This survey shows that Data NSW provides the type of environment and search capacity that developers want to engage with. Data NSW does need to provide more machine processable data however, and provide access to high value data that genuinely enables the development of real solutions to community problems.

[156] *Apps4NSW Optimisation: A market research presentation*, GFK, October 2015

f) Data NSW use and metrics

To provide further information about the implementation efficiency of Data NSW, this section examines use metrics and reporting statistics generated by the Data NSW portal.

A wide range of metrics about usage of the Data NSW portal are automatically generated through the native CKAN reporting tool with a Google analytics extension. [157] The CKAN site functionality also provides a breakdown of datasets by dataset characteristics. [158] In combination these metrics provide data on Data NSW:

- Site visits per month
- Number of organisations publishing data via Data NSW
- Most viewed datasets
- Most downloaded datasets
- Most edited datasets
- Dataset revisions per week
- Datasets published by organisations
- Datasets by format
- Datasets by license type

Total visits to Data NSW

The following table shows Data NSW site visits between May 2016 and May 2017. [159] These statistics show that Data NSW has had increasing usage over the last six months. Site usage statistics based on dataset use [160] show that this increase is driven by significant interest in the NSW Topographic Map [161] dataset in 2017:

Month	Total site visits
May 2016	5393
June 2016	5096
July 2016	4762

[157] Site statistics are captured at https://data.nsw.gov.au/data/stats#summary
[158] Dataset characteristics are accessible via https://data.nsw.gov.au/data/dataset
[159] Monthly site visit statistics are accessible via https://data.nsw.gov.au/data/site-usage
[160] https://data.nsw.gov.au/data/site-usage/dataset
[161] https://data.nsw.gov.au/data/dataset/nsw-topographic-map

August 2016	7009
September 2016	5510
October 2016	6383
November 2016	5858
December 2016	5667
January 2017	5246
February 2017	7084
March 2017	8675
April 2017	6150
May 2017	9954

Organisations publishing data on Data NSW

Currently 57 government organisations make data directly available via Data NSW.[162] A full list of the 57 government organisations making data available through Data NSW is listed at Appendix 3. Additionally, access to hundreds of other government organisations' datasets is provided through federation with data.gov.au. This federation means that keyword searches on Data NSW will also retrieve relevant datasets on federated portals.

Government organisations with the most viewed datasets

The 25 NSW government organisations with the most dataset views on Data NSW are listed below. Note that the agencies marked with asterisks have their own data portals or web links to interactive data visualisations, interactive maps, dashboards, reporting tools and RSS feeds, including the NSW Topographic Map, [163] *Mapping the Budget 2016-17*, [164] Fuel Check, [165] NSW Cancer Statistics [166] and NSW Parliament RSS feeds. [167] The dataset views in the table below do not represent activity on these portals or sites, only the specific views that pass through Data NSW. The actual dataset views, interactions, transactions or downloads for these organisations are therefore likely to be much higher when combined with their site or portal specific reporting. The number of agencies with asterisks in the list below shows the significant investment that many agencies have made to make data available, accessible and useable for community consumers.

[162] https://data.nsw.gov.au/data/site-usage/publisher
[163] http://maps.six.nsw.gov.au/arcgis/rest/services/public/NSW_Topo_Map/MapServer
[164] See http://myinfrastructure.planning.nsw.gov.au/
[165] https://fuelcheck.nsw.gov.au/App/
[166] http://www.statistics.cancerinstitute.org.au/
[167] https://www.parliament.nsw.gov.au/rss/Pages/rss-feeds.aspx

Publisher	Dataset views via Data NSW (October 2014 – June 2017)
Department of Finance, Services and Innovation *	31939
Department of Planning and Environment *	15515
Transport for NSW *	13502
Department of Education *	7078
Roads and Maritime Services *	6927
Fair Trading *	5475
Office of Environment and Heritage *	5385
Office of State Revenue	4701
NSW Bureau of Crime Statistics and Research *	4161
Family and Community Services *	2023
Bureau of Health Information *	1844
Ausgrid *	1816
State Records Authority *	1810
DPI – Water *	1700
Registry of Births, Deaths and Marriages *	1680
Public Works *	1517
Essential Energy *	1490
NSW Health *	1333
Anti-Discrimination Board	1300
Ambulance Service of NSW	773
City of Ryde *	762
NSW Rural Fire Service *	679
NSW Electoral Commission *	585

Department of Industry *	525
Parliament of NSW *	488

Most edited datasets, dataset revisions and PDF datasets

The most edited datasets list [168] and statistics for dataset revisions per week [169] show that only a small proportion of the datasets made available via Data NSW are routinely revised or updated with current data. While a portion of the Data NSW datasets are made available through APIs or web services which continually provide access to real-time data, many of the datasets on Data NSW are static and a significant proportion are available as non-machine processable PDF documents. Of the resources available on Data NSW, 166 are PDF documents. Some of these are human readable copies of machine processable data, but more than half of these datasets are only available in PDF form. [170]

The small proportion of datasets routinely revised or updated, and the significant proportion of datasets that are not machine processable limits the reuse potential of these datasets, by constraining their accuracy, currency, accessibility and useability.

Most downloaded datasets

While many datasets are viewed, the downloading of a dataset is a significant action which signifies an intent to engage with or to use a dataset in some way.

The list of the twenty most downloaded datasets via Data NSW October 2014 to June 2017 [171] is as follows. Note, as with the dataset view numbers in the table above, the dataset download numbers below do not incorporate all downloads through agency portals or sites, only the specific downloads that pass through Data NSW.

Dataset name	Dataset downloads via Data NSW (October 2014 – June 2017)
NSW Topographic Map [172]	1367
Greater Sydney Regional Housing activity	1192
NSW Cadastre	974
NSW Globe	845

[168] https://data.nsw.gov.au/data/stats#most-edited
[169] https://data.nsw.gov.au/data/stats#dataset-revisions
[170] See https://data.nsw.gov.au/data/dataset?res_format=PDF&_res_format_limit=0
[171] List derived from https://data.nsw.gov.au/data/site-usage/dataset
[172] Published July 2016

Fuel Check	724
Public Transport – Realtime – Vehicle Positions	378
NSW Base Map	366
Human Services Data Hub NGO providers	334
Rent and Sales Tables	322
Train Data	317
NSW Property	301
NSW Government School Locations and Enrolments	292
NSW Natural Resource Atlas	277
RMS Live Traffic Map – Developer Options	258
Planning and Environment Spatial Data	253
NSW Survey Control Information Management System	251
NSW Imagery	248
Standard Instrument Local Environment Plan – Land Zoning	228
NSW Budget Paper	224
Procurement Spend Data – Financial Year 2013-15	210
Speeding and Red Light Camera Offences	201
Baby Names Explorer	198

Learnings from the metrics analysis about the appropriateness of the Data NSW implementation:

- A significant proportion of NSW government agencies are not sharing data via Data NSW
- Publishing high value datasets (such as the NSW Topographic Map) can drive significant interest in and use of government data
- Publishing data in machine readable formats can greater facilitate its use and re-use

5.1.2 Are users being reached as intended?

a) List of agencies using Data NSW

The following agencies are using Data NSW.

Office of Environment and Heritage	Department of Education
Department of Finance, Services and Innovation	Bureau of Health Information
Environment Protection Authority	Transport for NSW
NSW Bureau of Crime Statistics and Research	State Archives and Records Authority
Fair Trading	City of Ryde
Anti-Discrimination Board	Department of Industry
Department of Planning and Environment	Office of State Revenue
Registry of Births, Deaths and Marriages	Housing NSW
Family and Community Services	Roads and Maritime Services
Public Works	Ambulance Service of NSW
Department of Primary Industry - Water	Macquarie University
Women NSW	Parliament of NSW
NSW Health	Juvenile Justice
Ausgrid	City of Sydney
Hornsby Shire Council	NSW Electoral Commission
NSW Rural Fire Service	Australian Bureau of Statistics
Australian Sports Commission	Tourism Research Australia
Murray-Darling Basin Commission	State Library of NSW

Sydney Water	Destination NSW
NSW Food Authority	Information and Privacy Commission
Property NSW	The Royal Botanic Gardens and Domain Trust
Essential Energy	Workcover NSW
Public Service Commission	Lismore City Council
Geographical Names Board of NSW	Australian Museum
Fire and Rescue NSW	Taronga Zoo
Hawkesbury City Council	Lord Howe Island Board
Powerhouse Museum	NSW State Emergency Service
TCorp	Legal Aid NSW
The Treasury	

Forty six of these 59 organisations are NSW government agencies. Therefore a proportion of state government agencies are users of the portal but a significant number of agencies are not releasing open data via Data NSW.

b) Survey responses – are users being reached as intended?

Findings from the DFSI Data NSW Key Stakeholder survey [173] address the question of whether key users are being reached as intended.

- 60% of respondents reported that Data NSW had led to improved transparency and accountability in their organisation, 20% said it had not lead to improved transparency and accountability and 20% were unsure. One respondent commented, 'I think that we are heading in the right direction, but we still have a long way to go.'
- 20% reported that Data NSW had supported evidence-based policy development in their organisation, while 20% said it hadn't and 60% were unsure

[173] The 30 question *Key Stakeholder Survey* was sent to key data contacts in agencies that are significantly engaged with open data in NSW: OEH, Planning and Environment, Transport, State Archives and Records Authority, Information and Privacy Commissioner, Department of Primary Industries. Six responses were received. This survey had a small sample size, but an informed and engaged constituency that provided detailed responses. This survey was conducted in May 2017.

- 40% said Data NSW had supported innovative solutions and improved service delivery, while 20% said it hadn't and 40% were unsure
- When asked whether Data NSW had provided positive opportunities within the organisation, 20% said yes, 20% said no and 60% were unsure. One respondent commented, 'I think that the organisation feels it is a lot of work to provide open data.'

These ambivalent metrics indicate that some positive outcomes are being enabled through Data NSW, but that ongoing work is required to ensure key government users who will drive the ongoing use and viability of the platform are being reached as required.

5.1.3 How satisfied are users?

a) Results from user surveys and interviews

Both the Key Stakeholder and Data NSW User surveys contained questions about user satisfaction. Responses and specific comments from the questions designed to identify the extent to which users are satisfied with Data NSW are listed below.

Satisfaction with Data NSW as reported in the Key Stakeholder survey

- When asked to rate their personal satisfaction with Data NSW from 1 to 5 (1 being 'not at all satisfied' and 5 being 'extremely satisfied'), the weighted average response of all respondents was 2.8 (between the responses 'it's OK' and 'quite satisfied'). One respondent reported being 'extremely satisfied'.
- When asked to rate their organisation's satisfaction with Data NSW from 1 to 5 (1 being 'not at all satisfied' and 5 being 'extremely satisfied'), the weighted average response was 2, which equated to the response 'it's OK'. One respondent reported being 'not at all satisfied'.
- When asked whether Data NSW helped to drive digital awareness and digital transformation in their organisation, 60% of respondents said yes. Two additional comments were provided to this question:
    - 'To a certain extent, it was good to have it there to 'formalise' our release of annual agency GIPA data, made it feel less ad-hoc. We now think more about what can be released.'
    - 'It provided a good structure around the open data space. Worked to provide additional examples of how we could do better at it.'

- When asked ‘How useful and relevant do you find Data NSW in enabling organisations to release and access open data?’, 60% said ‘quite useful and relevant’ and 40% ‘very useful and relevant’.

When asked to rate how useful Data NSW was in helping their organisation to share information from 1 – 4 (1 being ‘not useful and relevant’ and 4 being ‘extremely useful and relevant’), the weighted average response was 1.8 (between ‘not useful and relevant’ and ‘quite useful and relevant’). One respondent commented: ‘I think it could play a much bigger role as a ‘trusted interchange’ for data.’

Satisfaction with Data NSW as reported in Data NSW User survey

Respondents to the Data NSW User Survey [174] reported the following in response to questions on their satisfaction with Data NSW:

- When asked ‘how satisfied do you think your organisation is with Data NSW, 31% of respondents reported ‘not at all satisfied’. 25% were ‘quite satisfied’, 31% were ‘satisfied’, 13% were very satisfied. One agency commented: ‘Some issues with the way organisations are displayed (there are currently 3 organisations for OEH: NSW Office of Environment and heritage (OEH) with 0 datasets; ‘Office of Environment and Heritage’ with 21 datasets; ‘Office of Environment and Heritage (OEH)’ with 2742 datasets.’

- When asked ‘Does your organisation have data needs that are not being met by Data NSW?’ agencies reported:
    - ‘Data governance, data quality, data quality assessments, advice on privacy.’
    - ‘We have some information that requires a special portal to access. If Data NSW had the ability to provide links to this it would help.’
    - ‘Access to key performance data visualisations ie trend graphs.’

- When asked ‘If Data NSW was not available, would this have an impact on your organisation?’, 20% of respondents said yes, 43% said no and 37% were unsure. Comments include:
    - ‘It would be a shame to lose this resource and it would be good to develop it more appropriately – but as it is not well known to this organisation, it would be of little obvious consequence if it became unavailable.’

[174] The 14 question *Data NSW User Survey* was distributed to information professionals working in NSW public sector, through Information Management Community of Expertise and Privacy Professionals Network in May 2017. These information professionals are aware of data and data strategies in their organisations, but are unlikely to be directly responsible for open data strategies. It was also promoted on the Data NSW blog and reached a number of Data NSW private sector users through this channel. A range of responses were received from the public and private sectors. This survey received 37 responses.

- ‘I’d have to look elsewhere to publish Mapping the NSW Budget.’
- ‘We have our own data portal. The creation of policy is good though.’
- ‘We would still make the information available, but it might be harder for members of the public and interested individuals to find it.’
- ‘Increase manual demand and processing of data requests.’

- When asked ‘Did the existence of Data NSW help to drive any digital transformation or digital awareness in your organisation?’, 28% said yes, 42% said no, 31% were unsure. Specific comments included:
  - ‘Probably in terms of better understanding of the value of data to many users.’
  - ‘Mainly related to the ease of access and, in some cases, to actual availability of the data. Previously we had to negotiate a maze of various approvals processes to get access to data that we knew was there and would be useful. It’s not second nature for staff to refer to Data NSW as a primary source of information for our projects.’
  - ‘My impression is that it has, as there is a lot more awareness of data and its use in my agency, as compared to other agencies in which I have worked.’
  - ‘Prompted our own open data policy and portal.’
  - ‘We are an information and research service working with data for local agencies to assist with area and human service planning. Data NSW is an essential basic component of our activities.’
  - ‘Our open government program got a push by the Executives because of this, so thank you!’

User survey responses to the questions designed to measure satisfaction are quite ambivalent. The responses indicate a degree of satisfaction and identify a range of positive outcomes that have been achieved, but they do also flag that a number of users are not satisfied with Data NSW, or are minimally satisfied.

The responses indicate that there is still a large amount of work to be done to promote the value of open data, to facilitate open data release, and to improve the functionality of Data NSW to streamline the data publication process.

5.1.4 Has Data NSW delivered value for money?

a) Program funding

The annual budget for Data NSW is very moderate at $30,000. This principally covers site hosting costs and a small amount of maintenance costs and arrangements provided by a Canberra-based IT company. Work to improve data maturity, quality and collaboration is undertaken as unbudgeted business as usual or project work within SARA and DFSI, alongside other work priorities.

Between the work performed by SARA and DFSI to support Data NSW, the FTE in maintenance and support (and associated cost) is currently 1.5 at a Band 9-10 rate.

A multi-year funding model should be developed with a mix of recurring funding for day to-day governance, operations and maintenance and one-off funding for developments such as new products and service features or to support updates and enhancements.

A future funding model should also align with future strategic directions for Data NSW that are not related to its technical environment. For example, if a future requirement for Data NSW is to increase the amount of data published on the site, funding may be allocated to train data publishers and to re-develop publication processes on the site. If data re-use is to be a key objective, funding may be used to train for data users, for engagement activities or for focussed challenges, competitions or events around specific problems or issues. Limiting funding to technical aspects of the portal constrain user-engagement or other events which are potentially more important to the sustainability and impact of the portal.

Data NSW is built on a CKAN platform which is affordable and open source. However CKAN's rich technical backend means it is hard for teams with limited technical resource to maintain and update. CKAN does enable significant functionality, interoperability and connection to a global open data user base, but it requires knowledgeable developer support for updates and changes which means even small site alterations can be costly. Therefore any budget allocations do need to maintain adequate support and development costs.

Much technically is possible on the CKAN platform however, and as the dominant open data portal globally, there are many examples and case studies to draw from. It has an API and can be used to provide rich data resources like visualisations. It is worth noting that for an open data portal to offer data visualisations, data typically needs to be stored on the portal as structured and machine-readable data. Storing data on the portal raises other resourcing implications and so it is important to consider the full dimension and implication of functionality decisions.

There are not many publicly accessible sources of open data portal annual costs. However the report, *How to make an open data portal sustainable*, interviewed twenty people involved in managing open data portals across Europe. When asked about the annual budgets for their portals, the portal owners who were able to provide an estimate stated that their annual budgets ranged from €90,000 to €700,000. Other portal owners reported that

their funding only supported basic day-to-day portal operations rather than developments, with comments such as: 'With just three of us, we have no capacity to progress it, only really [to keep] it ticking over'. [175]

In terms of generating revenue from an open data portal, the report strongly advises against charging for data that should be freely available as open data. It notes that activities such as training courses, service-level agreements for high-volume API access, or data analytics services can be provided on top of a portal to generate a revenue stream.

It also cautions however that in dynamic data environments, portals should not offer too many services as this could hamper business opportunities and innovations that third parties could develop using open data as a base. The report provides the following case study from Transport for London (TfL) to illustrate this downstream economic potential:

> TfL created and supported numerous APIs that provide transport data – such as timetables and live access to bus and train locations, and departure and arrival times – and convened an Open Data user group for developers to collaborate. This in turn led to the creation of travel information apps, such as CityMapper, which help users to plan journeys and avoid disruptions.
>
> Allowing apps such as these to emerge has created huge benefits for the data supplier, data users and service users. One study estimated that TfL had generated a return on investment of 58:1 from passengers using the service more efficiently. The Shakespeare Review estimated the value of this time saved by users to be £15-58 million each year.
>
> These benefits were possible because TfL did not use its position as data supplier to unfairly hamper data users downstream. Citymapper's general manager Omid Ashtari said, 'Citymapper was created [in the UK] because of the existence of Open Data. It's the essential backbone of what we're working on.' [176]

Despite their critical role, most Open Data portals were not set up with sustainability in mind. They were mostly created quickly as part of an emerging political movement without feasibility studies, business case, strategy or user research.

To move from setup to sustainability Data NSW needs to embed ongoing and adequate budget allocations into its operations, in order to best meet its goals and to adapt to the challenges it faces.

b) Assessment of resourcing requirements

Administration of Data NSW involves supporting the range of processes outlined below.

The State Archives and Records Authority provides:

[175] https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf, p30
[176] https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf, pp32-33

- Data NSW technical and daily operational support
- User registration and management, which, in accordance with the defined process for signing new publishers to Data NSW, involves granting authorised users permission to use the service and ensuring that unauthorised users cannot perform custodian activities,
- Project management for Data NSW development work
- Site federation work and coordination with other NSW data portals and state and national CKAN open data sites
- Contract liaison
- Vendor support liaison
- Helps agencies to upload data
- Advises on broken links and other relevant data quality or accessibility issues

The annual vendor support contract for platform support is approximately $30,000. For this cost the vendor support partner:

- Hosts the Data NSW site
- Resolves technical issues
- Provides 24/7 technical support
- Makes changes to structure of site pages
- Detecting, reporting and handling routine or unanticipated events
- Restoring operations after an interruption to the normal running of the service
- (For an additional cost) Provides site development services to improve Data NSW functionality and services

The Information team in the Department of Finance, Services and Innovation:

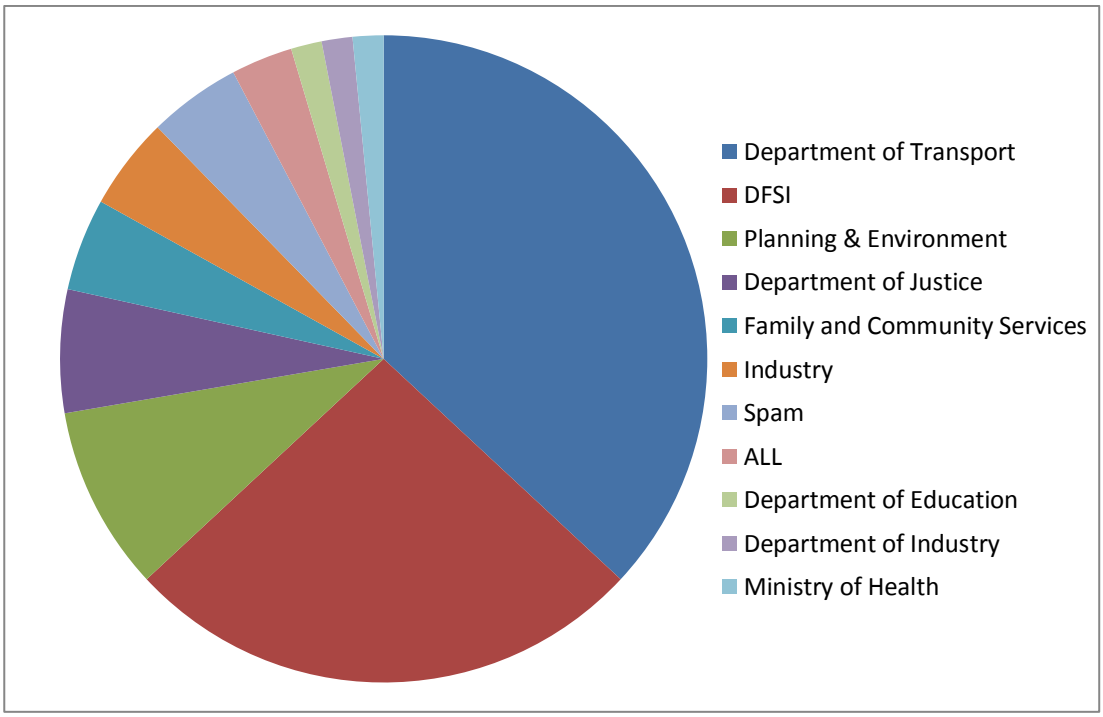
- Develops open data policies and procedures
- Works with agencies to build an open data culture and to implement open data policies and procedures
- Develops communications, blog updates and social media communications to support open data and use of Data NSW

- Works to increase usability and awareness of open data and Data NSW
- Reports on Data NSW activity

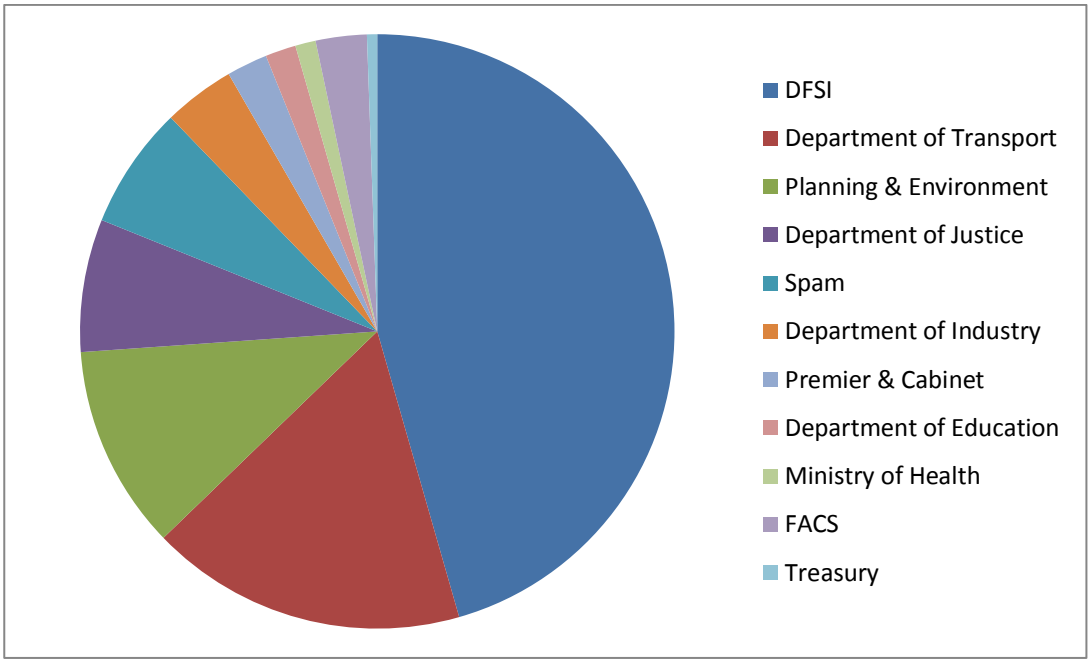
In terms of managing public enquiries made through the Data NSW portal, the Information team:

- Responds to public enquiries. Sixty five enquiries were received in 2014-15. One hundred and eighty were received in 2016. The pie charts below provide a list of the agencies that members of the public had data or dataset enquiries for.
- Refers requests for new data uploads to specific agency or to agencies likely to hold requested data
- Refers dataset quality or accessibility issues to relevant agency
- Maintains and updates agency contact information.

2014-15 Data NSW mailbox requests by target agency:



2016 Data NSW mailbox requests by target agency:



Resourcing for public enquiries

There is an ongoing issue with providing adequate resourcing for public enquiries.

Responding to end-user requests places a high demand on resources and time is frequently consumed responding to requests that cannot be met, or requests for specific statistics, rather than data.

Problems with datasets are reported to the Data NSW mailbox and are then escalated to appropriate data custodians. A direct custodian contact point for datasets could minimise this double-handling. An updateable register on Data NSW of bugs, datasets with broken links, missing resources, or upcoming feature requests under development could also streamline issue management.

Standards of service or service level agreements have not been defined for public or agency requests to Data NSW. No statements are provided to manage public expectations, and agencies are unaware of resourcing requirements to support public request management. Standards of service for the Data NSW mailbox should be defined and publicly posted on the site.

Understanding users

Currently there is no overview of the typical profile of Data NSW users. It is difficult to know who users are, and therefore challenging to know how to best meet their needs. Users also do not have to register to use Data NSW and can therefore remain anonymous. Data NSW users are therefore not an easy group to contact, survey or request information from about their experiences with and use of the site.

It would be useful to be able to measure reuse of data, and to understand why and how users are making use of data, but capturing these metrics is challenging. The next iteration of the *Sustainability of open data portals* report will explore the development of automated metrics for measuring the use and impact of open data. This report is due at the end of 2017.

Monitoring data quality

Currently there are no automated processes on Data NSW or other open data portal to ensure data quality, other than ensuring key mandatory metadata fields are completed during the publication process. On Data NSW, no automated processes are in place to check for errors, or to alert publishers when their data is out of date.

The European *Sustainability of open data portals* report identified that what the majority of portal owners surveyed wanted to measure about their data was quality and re-use. As the Norwegian portal owner put it, 'sustainability comes down to data quality'. [177] Currently Data NSW has on measures in place to monitor data quality.

Metrics on impact

Data NSW reporting metrics primarily focus on quantity measurements alone. These however are insufficient to measure impact. Number of datasets available is frequently used as a benchmark measure for the site's impact and success. However it is difficult to know exactly what constitutes one dataset and therefore this figure is often unreliable. This figure also reveals nothing of dataset use and impact. It also does not assess important indicators such as:

- dataset quality
- whether documentation published alongside datasets is accurate and comprehensive
- data published in open, machine-readable formats
- data published with open licenses.

Quantity benchmarks also drive the publication of quick wins or low value datasets and the large-scale publication of poor quality data can diminish the impact of open data initiatives.

Quantity metrics are however very available and are easy to automate and measure. European research has demonstrated the challenge of trying to use automated metrics to measure the impact of open data. [178] A 2013 research initiative concluded that the more meaningful a construct appears, the harder it is to measure. [179] For example, highly available and measurable metrics include:

[177] https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf, p50
[178] http://opendatamonitor.eu/
[179] http://project.opendatamonitor.eu/wp-content/uploads/deliverable/OpenDataMonitor_611988_D2.3-Best-practice-visualisation,-dashboard-and-key-figures-report.pdf

- Conformance to a specific standard/s
- Timeliness
- Machine readability
- Number of datasets.

Much more difficult to measure metrics include:

- Impact
- Quality
- Completeness
- Documentation.[180]

It is therefore difficult to evaluate the true impact of Data NSW, and the metrics that are more easily available do not provide the full picture about the use of the site.

Data retention and removal policies

Data NSW needs to consider the provisions it needs to have in place to archive data or to respond when data disappears. Currently there are no policies in place to govern these important operations. Data that third parties rely on can therefore be removed without consultation, and there are no requirements for data longevity made on the site.

Key learnings on whether Data NSW delivers value for money

- Data NSW definitely delivers value for money. It runs on minimal costs but has still released very substantial amounts of data to the public and still drives active data publishing and data download activities every month.
- The current Data NSW budget is allocated to the technical environment
- A future funding model should also align with future strategic directions for Data NSW that are not related to its technical environment
- Data NSW needs to embed ongoing and adequate budget allocations into its operations, in order to best meet its goals and to adapt to the challenges it faces
- Data NSW needs data retention and removal policies to ensure third parties can rely on data stability and longevity

[180] https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf, p58

c) Key Stakeholder survey on value of Data NSW

Key findings from the Key Stakeholder survey [181] on the perceived value of Data NSW to agencies are as follows.

- 60% of respondents reported that Data NSW had led to improved transparency and accountability in their organisation, 20% said it had not lead to improved transparency and accountability and 20% were unsure. One respondent commented, 'I think that we are heading in the right direction, but we still have a long way to go.'
- 20% reported that Data NSW had supported evidence-based policy development in their organisation, while 20% said it hadn't and 60% were unsure
- 40% said Data NSW had supported innovative solutions and improved service delivery, while 20% said it hadn't and 40% were unsure
- When asked whether Data NSW had provided positive opportunities within the organisation, 20% said yes, 20% said no and 60% were unsure. One respondent commented, 'I think that the organisation feels it is a lot of work to provide open data.'
- When asked whether Data NSW has empowered agency clients and customers, 20% said yes, 40% said no and 40% were unsure
- When asked if Data NSW had addressed other needs in their organisation respondents reported:
    - 'We have been able to use the platform to access data to participate and host open data challenges eg apps4NSW and GovHack'.
    - 'Promoting open data in NSW'
    - 'Providing connections within different clusters, to learn how they overcome these problems'.
- When asked if Data NSW has helped agencies to solve problems, agencies reported:
    - 'It gives us a framework/standard approach for any data release, rather than inventing our own approach.'
    - 'We strongly support the expansion of Data NSW to provide more advice/support to agencies in releasing and sharing data'.

[181] The 30 question *Key Stakeholder Survey* was sent to key data contacts in agencies that are significantly engaged with open data in NSW: OEH, Planning and Environment, Transport, State Archives and Records Authority, Information and Privacy Commissioner, Department of Primary Industries. Six responses were received. This survey had a small sample size, but an informed and engaged constituency that provided detailed responses. This survey was conducted in May 2017.

d) Data NSW user survey

Key findings from the *Data NSW user survey* [182] are as follows.

Perceived value of Data NSW

- Respondents were asked to indicate on a sliding scale from 0 – 100, how much value they think their organisation sees value in Data NSW. The weighted average response was 57%, indicating that agencies see a slightly above average value in Data NSW.
- Respondents were asked to explain why they thought Data NSW provided value and were able to select more than one response. 87% of respondents said Data NSW offered value to government. 83% said it provided value to the community. 67% said it provided value to industry. 67% said it offered value to their organisation. 53% said it offered value to innovation. Comments included:
    - 'Its value continues to mature but I think you need to go harder, faster and more definitive.'
    - 'Data does provide value, but it depends on there being a rich variety of data, and also on people knowing how to use it. In my view most people in government don't really know how to use data. They get excited about it, or fear it, but don't have a deep understanding of what is required to really use it well.'
    - 'For it to be effective it needs all agency and all government data available. It can be quite frustrating to know data exists but one cannot find it or the contact to ask.'
    - 'It is great to have one place where data can be released and where people can request access to other information that could be useful to research, innovation or otherwise in solving issues.'
    - 'My organisation has its own open data portal but Data NSW is more accessible.'
    - 'It is great that all the data is available at a single point of location, easily searchable and well organised.'

[182] The 14 question *Data NSW User Survey* was distributed to information professionals working in NSW public sector, through Information Management Community of Expertise and Privacy Professionals Network. These information professionals are aware of data and data strategies in their organisations, but are unlikely to be directly responsible for open data strategies. It was also promoted on the Data NSW blog and reached a number of Data NSW private sector users through this channel. A range of responses were received from the public and private sectors. This survey received 37 responses. The survey was conducted in May 2017.

- ‘Data NSW potentially at least, offers value to all sectors. However ease of use and management with the existing system are issues – until we had the automatic harvesting working, it was an unrealistic overhead to upload each dataset separately to Data NSW and keep metadata and resources up to date in Data NSW as well as in OEH corporate systems. Our Open Data Portal was developed in CKAN specifically so that uploads to Data NSW could be automated, and it is very important that this ability is maintained in any future redevelopment of Data NSW.’

- A private sector respondent to the survey provided an interesting comment on the value of Data NSW: ‘Private sector organisation not very aware of Data NSW and potential value to industry. Not much data of value to industry currently available on Data NSW. Organisation would mainly think of Data NSW being a vehicle for govt transparency.’

Respondents to the *Data NSW user survey* were also given the opportunity to explain why they thought Data NSW did not provide value. Respondents could provide multiple responses.

58% of the 37 respondents said that Data NSW’s value was not understood across the community. 47% said its value was not understood in their organisation. 37% said its value was not understood across industry. 21% said their organisations had their own open data portals and did not need Data NSW. Specific comments that address the question of whether users are being reached as intended included:

- I think this initiative has been swallowed up in a lot of other information and data initiatives occurring at the similar time. I think as an agency we’re also a little reserved about putting data up that’s relevant and in good condition.
- Data NSW has been poorly resourced for some time and is not well known to ‘new comers’ in the data space. It has been outpaced by developments in other jurisdictions or agency/subject specific portals. Separate portals means disjointed search and access. Agencies do not proactively feed Data NSW.
- For it to be effective it needs all agency and all government data available. It can be quite frustrating to know data exists but one cannot find it or the contact to ask.
- My organisation has too many other reforms occurring for open data to have any chance of getting cut through in the general flow of information with any prominence. It is very difficult to see this as being likely to change in the foreseeable future.

Key lessons to improve value of Data NSW:

- A future multi-year funding model, based on strategic directions for Data NSW needs to be developed

- Data custodian identity management needs to be well managed and an accurate list of current contacts maintained to facilitate ongoing dataset support and customer enquiry management
- Standards of service should be developed for the Data NSW mailbox
- More frequent user surveys should be undertaken
- Metrics on impact should be considered
- Data retention and removal policies should be developed to ensure appropriate ongoing accessibility of high value datasets

6. Key findings: Outcomes

6.1 Effectiveness

6.1.1 To what extent has the Data NSW achieved its objectives?

a) Objectives of Data NSW

The objective of Data NSW is to make more government data available to the public and increase the flow of data shared internally within government.

The shorter term expectations of Data NSW was to:

- improve the transparency and accountability of the NSW Government
- support evidence-based policy development
- enable innovative solutions and service delivery
- improve government services
- empower citizens
- create positive opportunities
- enable government to solve key problems.

b) Metrics and survey data

The current metrics available on Data NSW demonstrate that large volumes of data have been release and that regular downloads are occurring via the platform. This large scale and routine public access of government data is evidence that Data NSW has improved the transparency and accountability of the NSW Government and how citizens access data.

Key Stakeholder survey responses however indicated that Data NSW has not yet comprehensively contributed to the broader objectives of contributing to evidence-based policy and innovative service delivery.

It does need to be noted that this survey had a small sample size of six, but as the respondents were all key stakeholders of Data NSW and are open data advocates in their organisation, their responses have credibility and value.

The following tables list the statements proposed in the survey and open data advocate respondent answers to them:

Data NSW has supported evidence-based policy development in my organisation		
Yes: 20%	No: 20%	Unsure: 60%

Data NSW has supported innovative solutions and improved service delivery in my organisation		
Yes: 40%	No: 20%	Unsure: 40%

Data NSW has provided positive opportunities in my organisation		
Yes: 20%	No: 20%	Unsure: 60%

Data NSW has empowered agency clients and customers		
Yes: 20%	No: 40%	Unsure: 40%

A second survey was distributed to Data NSW users, generally information professionals working in NSW public sector, with 37 responses received. These information professionals are aware of data usage and data strategies in their organisations, but are unlikely to be directly responsible for applying open data strategies in their workplace.

The following tables list the statements proposed in the survey and the information professional respondent answers to them:

If Data NSW was not available, would this have an impact on your organisation?		
Yes: 20%	No: 43%	Unsure: 37%

Did the existence of Data NSW help to drive any digital transformation or digital awareness in your organisation?		
Yes: 28%	No: 42%	Unsure: 31%

c) Areas for improvements

The survey data shows some areas of achievement and areas that can be built on. A 'yes' response was provided for each outcome measurement question which evidences that, to a certain extent, Data NSW's outcomes have been achieved in organisations.

There are clear areas for improvement however in expanding the reach of these outcomes across all agencies.

6.1.2 Were there any unintended impacts (positive or negative)?

a) Results from user surveys and interviews

Survey responses indicated that there are positive impacts of Data NSW, both currently and potentially in the future.

For example, one respondent suggested that Data NSW should play a larger role as a 'trusted interchange' for data. This indicates that Data NSW has a trusted reputation and suggests that there is a growing government appetite for the safe and secure sharing of data.

Another positive impact identified in survey is that Data NSW has provide a standard approach for any data release, which has saved agencies from having to invent and invest in their own approach.

On the other hand, an unintentional impact of Data NSW is that it has contributed to agencies' concerns and anxieties about data sharing. One respondent reported that 'managing the risk of data misuse' was a key challenge with open data, and numerous survey responses pointed to an ongoing concern with open public data. One respondent questioned the rationale behind open data by saying 'Don't understand why Government organisations would want to spend time and energy to hand over data for free to third party developers who then go and make money off it.' This suggests that there are conflicting views around the Open Data Policy principle of "free where appropriate".

Another negative impact relates to confusion between the role of Data NSW and other NSW government data portals. Various survey respondents made comments like 'Get a clear direction/objective of Data NSW' and 'Clearly message why do we need to use Data NSW vs not or other entities'. Other responses counter this however by saying 'My organisation has its own open data portal but Data NSW is more accessible' and 'It is great that all the data is available at a single point of location, easily searchable and well organised.'

Data NSW exists as an aggregator and a means of centralising search for public and private sector users. It does not replace any other portal but it intended to increase connectedness and access potential. The integration back-end of Data NSW should be streamlined to automate as much as possible connection between data portals, and to

ensure that data custodians do not have to manually register their datasets on their own portals as well as Data NSW.

b) Areas for improvements

It is clear that Data NSW needs to better articulate and outline its role, purpose and outcomes to both the public and NSW Government agencies. Improving the messaging around the benefits of the portal may facilitate greater usage and participation throughout NSW Government.

It is also clear that there is a need for various technical improvements to make the integration of Data NSW with other data portals as seamless as possible. When Data NSW was first launched, no NSW Government agency had a data portal. The Office of Environment and Heritage, Transport for NSW and Education, among others, have their own portals with various functionalities. While Data NSW does federate and link to these sites, there are opportunities to improve this process.

7. Learnings

The Data NSW program evaluation has provided a wide range of learnings that can be grouped under the following categories:

- User experience
- Quality environment
- Culture and collaboration
- Technical environment
- Budget
- Metrics

This section provides a summary of the key learnings across each of these areas.

7.1 User experience

Key lessons relating to user experience are as follows.

7.1.1 Focus on interface design

The evaluation applied the data portal gap analysis framework developed by Renata Máchová and Martin Lnenicka to assess the quality dimensions and general characteristics of Data NSW. Data NSW faired strongly against this assessments (see results at section 5.1.1d). However, the evaluation also identified a number of areas for improving the interface design of Data NSW.

Data NSW should present a simple and clean interface, with a clearly visible dataset search tool function. The homepage should present less script to provide enhanced user orientation.

Open data interfaces should focus on key content, search, navigation and accessing information. These core functions should use consistent heading sizes, font choices, colouring, button styles, spacing, design elements, illustration styles to enhance user experience of the site.

Open data interfaces should make the data publishing process as simple and efficient as possible for data custodians and publishers. Data.gov.uk and DataSF provide excellent examples to follow.

7.1.2 Enable public feedback and contact with data custodians

Open data frameworks should enable and facilitate an open dialog between users and data custodians. Each dataset should provide contact details of data custodians or a single

data@ email address for each agency, rather than the current under resourced centralised inbox which adds an additional administrative layer.

7.1.3 Enable users to better understand data

Features should be consider that further enable citizens to track and better understand the source of open data, and that enable improved discovery of data insights and related underlying data. ‘Tips for users’ as well as user centred design can help the public to be much more engaged with data and aware of how they can use it.

7.2 Quality environment

7.2.1 Develop a strategic and coordinated approach to data quality

The program evaluation has highlighted the critical importance of the quality environment.

The evaluation has shown that quality takes many interrelated forms including:

- Governance
- Baseline quality criteria
- Guiding principles
- Metadata
- Licensing
- Vocabularies and code lists
- Registers.

For a high quality data environment, work on all these pieces and incorporation in open data frameworks should be considered.

7.2.2 Standardisation is a fundamental enabler of quality and data use potential

The United Kingdom provides strong examples of the standardisation necessary to build a strong open data culture. Standardisation enables data reuse, tool design and data consistency and reliability. Key examples also come from the World Bank environment of the benefits of significant investment in data consistency and quality.

To facilitate better sharing, it is important that data ownership and privacy requirements are clearly defined.

7.2.3 Make open data machine readable data

A large proportion of NSW open data is currently not machine readable. Agencies should be supported to make more of their data available in machine readable formats to allow for its greater reuse potential.

7.2.4 Outline data retention and removal policies

Data NSW needs data retention and removal policies to ensure third parties can rely on data stability and longevity.

7.2.5 Data custodian management needs to be rigorous

Data custodian identity management needs to be well managed and an accurate list of current contacts maintained to facilitate ongoing dataset support and customer enquiry management. This also maintains the integrity, useability and ongoing accessibility of datasets.

7.3 Culture and collaboration

7.3.1 Build support for open data across government

A significant proportion of NSW government agencies are not sharing data via Data NSW.

Support for open data exists but genuine culture change is still necessary to fully enable open data in NSW government. Survey results reveal a disconnect between stated open data principles and genuine open data practice.

Some key users within and beyond government are being reached and positive outcomes are being enabled through Data NSW, but ongoing work is required to ensure key government users who will drive the ongoing use and viability of the platform are being reached as required.

The moderate levels of satisfaction currently reported on in the user surveys could be improved by building greater support and by augmenting the site's technical features.

The Information Commissioner provides current high level support as NSW Open Data Advocate. This support should be leveraged to enable open data and drive organisational open data support at the executive and business levels. This will help to raise awareness of open data and will help to build strong governmental drivers for open data.

The financial use cases for open data are very compelling and should be used to drive support for open data across government.

Support for data sharing should also be encouraged across government.

It is also important for NSW to operate as part of a national data framework and to build data frameworks that are interoperable with federal, state and local government frameworks. This level of interoperability delivers the greatest value and return on investment to the community.

7.3.2 Enable agencies to open data

The NSW Government should develop a toolkit to consolidate its existing open data guidance material and to provide agencies with clear steps on how to publish open data. This will minimise unnecessary queries to the Data NSW website. Numerous excellent examples exist in the federal government, UK government, World Bank, San Francisco and Washington DC open data environments.

Agencies should be helped to understand the benefits of releasing primary and timely data.

7.3.3 Help users to understand and use open data

Advice should be developed to help citizens understand what open data is and how it can be used. Regular surveys or consultation should also occur with the public to identify that their open data needs are and to determine how these needs can be met. The community should also be engaged in dataset release prioritisation and design. Publicly listing all government datasets and committing to ongoing forward release as shown by the data publication schedules on DataSF can also help the community to understand and use open data.

7.3.4 Release the data that the community wants and needs

Open data strategies need to provide access to high value data that genuinely enables the development of real solutions to community problems.

Open data programs should engage directly with the community to identify the data that they want released.

7.3.5 Engage more with social media

Promote and advertise data releases in ways that engage the community and that share data-driven insights with the widest possible audience to build a strong and collaborative open data culture. Data.gov.uk and DataSF provide excellent examples of engaged blogging practice and the @worldbankdata Twitter account is an exemplar in micro-blogging practice.

7.3.6 Understand that anonymization and de-identification are drivers for open data

To enable open data release, detailed and consistent guidance needs to be in place on anonymisation and de-identification. This advice should provide practical case studies,

examples and detailed advice for agencies that will assist them to enable open data release. Extensive but practical models for this advice have been published in the UK.

7.3.7 Employ agile work practices to build a dynamic culture

By taking an agile, collaborative and iterative approach to its work, DataSF has built an engaged culture, has been quick to learn from its experiences and is able to innovate the San Francisco data environment.

7.3.8 Promote awareness of emerging trends and open data's place within them

A key outcome of the Internet of Things assessment is an understanding of the economic potential of platforms like Data NSW that make stable, valuable, standardised and guaranteed sets of government data available for business and community re-use.

Awareness of this form of critical infrastructure can build support for and commitment to open data, which will then increase the economic and business benefit of open data platforms.

7.3.9 Facilitate agency collaboration to share knowledge

Agencies struggle to find the resources and time to make data open. Sharing knowledge, expertise, contract or service requirements, lessons learned, resources, skill sets etc is an important component of reducing the time and resourcing needed to support and publish open data. Key open data stakeholders have significant insights to share on open data enablers and inhibitors that can benefit all open data initiatives.

7.3.10 Understand and support the amount of change agencies are experiencing

Agencies are experiencing extensive transformational change. Open data initiatives as integrated parts of other tools or drivers, such as the Digital Strategy or standard procurement arrangements, to minimise impact and maximise benefit for agencies.

7.3.11 Use data to solve real problems

Data must be used to solve problems and sharing these stories will help to build a strong open data culture.

7.4 Technical environment

7.4.1 Support for data visualisations

Metadata and data storage layers should be designed to enable data visualisation to be used as much as possible, so that users can easily and quickly get an

understanding about the dataset, and decide whether it is worth continuing with a more detailed analysis.

7.4.2 Enable open data by design

The data NSW government holds is often locked into inflexible IT systems and retrieving the data can be a costly exercise requiring detailed business case or contractual amendment. Better, flexible and more intelligent IT systems supported by transparent contracts are essential. Future IT contracts must allow for easy and uninterrupted access to data held on the behalf of NSW government agencies.

Contractual agreements around projects, services, grants, research and a wide range of government deliverables should also outline clear requirements for open data outcomes as a component of contract delivery.

7.4.3 Utilise APIs

API-enabled search functionality should be used to augment the current search functionality of Data NSW.

API access to high value datasets should also be enabled wherever possible.

7.4.4 Maintain interoperability

The work performed by NSW government agencies to ensure interoperability of their open source data portals with the open source NSW data portal needs to be maintained. The European portal research also shows that enabling data accessibility through interconnected and interoperable data portals is an important component of the user experience and important for maximising data availability.

7.5 Budget

7.5.1 Data NSW has delivered value for money

Data NSW runs on minimal costs but has still released very substantial amounts of data to the public and still drives active data publishing and data download activities every month.

7.5.2 Open data budgets need to support strategic directions

Develop funding strategies and budgets that support the portal's planned and future needs and directions, not just technical and maintenance requirements.

7.5.3 Budget allocations need to be stable and ongoing

Data NSW needs to embed ongoing and adequate budget allocations into its operations, in order to best meet its goals, provide a stable resource for innovation and to adapt to the challenges it faces.

7.6 Metrics

7.6.1 Work to try and measure impact of open data initiatives

Open data is meaningful if it is used. Efforts should therefore be made to develop meaningful measures and indicators that try to track the use and impact of open data.

8. Recommendations

The following key recommendations are made based on the research and learnings represented in this program evaluation. Implementing these recommendations will help Data NSW to further achieve its objectives, increase user satisfaction and deliver value for money.

1. Identify the work required to improve the Data NSW user experience, focussing on interface requirements, streamlined approaches to data publication, improved public contact with data custodians and tools to enable users to better understand data.

2. Identify the necessary components of a data quality framework in NSW, and identify best practice national and international resources that could be adapted or adopted as part of this framework.

3. Work with NSW government agencies to identify the drivers and inhibitors of open data release in NSW.

4. Develop strategies based on identified data release drivers and inhibitors to improve open data release in NSW government.

5. Utilise social media to engage the community and public sector in open data.

6. Undertake research into the open data frameworks, metadata standards and quality tools required to support further data visualisation, spatial enablement and the increased development of APIs.

7. Investigate options for sustainable and ongoing funding options for Data NSW.

9. Appendix

This appendix provides the following stakeholder engagement tools used in the evaluation:

- Data NSW User Survey
- Data NSW Key Stakeholder Survey

9.1 Data NSW User Survey questions

1. Are staff in your organisation aware of Data NSW?
    a. Yes, many staff are aware of Data NSW
    b. Yes, some staff are aware of Data NSW
    c. No, only staff who work with data are aware of Data NSW
    d. No, hardly any staff are aware of Data NSW
2. Do you think people in your organisation understand the role and purpose of Data NSW?
    a. Yes
    b. No
    c. Unsure
3. Which areas of your organisation do you think have a good understanding of the role and purpose of Data NSW? (Please select all that apply.)
    a. The majority of staff understand the role and purpose of Data NSW
    b. Executive understand the role and purpose of Data NSW
    c. Data stewards understand the role and purpose of Data NSW
    d. Business areas understand the role and purpose of Data NSW
    e. My business area understands the role and purpose of Data NSW
    f. Unsure

4. Is your organisation committed to the release of open data?

   a. Yes, and this commitment is specified in an open data policy

   b. Yes, and this commitment is specified in a policy other than an open data policy

   c. Yes, but this commitment is not formally specified in any policy

   d. No

5. Do you think your organisation sees value in Data NSW? Please use slide to indicate.

6. If you believe Data NSW provides some value, please tell us why. Select all that apply.

   a. If offers value to the community

   b. It offers value to industry

   c. It offers value to government

   d. It offers value to innovation

   e. It offers value to our organisation

7. If you believe Data NSW does not provide value, please tell us why. Select all that apply.

   a. Its role is not understood in my organisation

   b. Its role is not understood across industry

   c. Its role is not understood across the community

   d. My organisation has its own open data program and does not need Data NSW

   e. My organisation does not see value in Data NSW

8. If your organisation has uploaded datasets to Data NSW, when was this?

   a. In the last 6 months

   b. In the last 12 months

c. In the last 12 - 24 months

d. In the last 3 - 5 years

9. If your organisation has not uploaded datasets to Data NSW, why is this?

   a. We don't have any appropriate data

   b. We don't have appropriate resourcing

   c. We don't have appropriate skills

   d. We have privacy and safeguarding concerns

   e. Our corporate culture makes open data challenging

   f. Uploading data to Data NSW is difficult

   g. We are currently working to get some datasets ready for publication

10. If you have uploaded datasets, do you know how they have been used by government, industry or the public?

    a. Yes

    b. No

    c. Unsure

11. How satisfied do you think your organisation is with Data NSW?

    a. Not at all satisfied

    b. Quite satisfied

    c. Satisfied

    d. Very satisfied

    e. They love it

12. Does your organisation have data needs that are not being met by Data NSW? If so, please provide details.

13. If Data NSW was not available, would this have an impact on your organisation?

    a. Yes

    b. No

    c. Unsure

14. Data NSW was launched in 2009 and was the first state-based open data portal in the world. Did the existence of Data NSW help to drive any digital transformation or digital awareness in your organisation?

    a. Yes

    b. No

    c. Unsure

9.2 Data NSW Key Stakeholder Survey

Background

1) What agency do you work in?
2) What has been your involvement with data.NSW?

Interview questions

3) We are interested in hearing your views on the extent to which agency staff are aware of data.NSW and are using it for its intended purpose:
    a. To your knowledge, what is the level of awareness of data.NSW in your agency?
    b. Do you feel staff in your agency understand the purpose of data.NSW?
    c. Have you or anyone else in your agency used Data.NSW to upload datasets for public access? (If yes, how often?)
    d. Have you or anyone else in your agency used Data.NSW to search for and use publicly available datasets? (if yes, for what purpose?)
    e. Have you or anyone else in your agency registered datasets on the Information Asset Register? (If yes, how often)
    f. Have you or anyone else in your agency used the Information Asset Register to search for restricted datasets held by other agencies? (If yes, for what purpose)?
    g. Have you or anyone else in your agency used Data.NSW for any other purposes? (If yes, what purposes e.g., apps showcase, twitter feed, FAQ section etc)
    h. Does your agency have an Open Data Strategy? (If yes, does it refer to data.NSW?)
    i. Does your agency have an Agency Information Guide? (If yes, does it refer to data.NSW?)

4) We are interested in understanding how satisfied users are with data.NSW and how it could be improved:
    a. How satisfied are you with data.NSW?
    b. How satisfied do you feel other members of your agency are with data.NSW?

c. Have you or anyone else in your agency encountered issues with data.NSW? If yes:
    i. what issues have been encountered?
    ii. were these issues resolved?
    iii. did you receive adequate support to overcome these issues?
d. Do you have any recommendations on how data.NSW could be improved?

5) We are interested in hearing your views on whether data.NSW addresses an identified need:
    a. Do you think that data.NSW addresses an identified need?
    b. How useful and relevant do you find data.NSW in enabling agencies to release and access open data?
        i. What are the main challenges that your agency has with open data more generally?
        ii. How do you think data.NSW could be improved to address these challenges?
    c. How useful and relevant do you find data.NSW is in enabling your agency to share data?
        i. What are the main challenges that your agency has with shared data more generally?
        ii. How do you think data.NSW could be improved to address these challenges?

6) We are interested in assessing whether data.NSW has achieved its objectives:
    a. Do you feel that data.NSW has led to improved transparency and accountability within your agency? (If yes, how and can you provide any specific examples)?
    b. Do you think that data.NSW has supported evidence-based policy development within your agency? (If yes, how and can you provide any specific examples)?
    c. Do you think that data.NSW has supported innovative solutions and improved service delivery in your agency? (If yes, how and can you provide any specific examples)?
    d. Do you feel that data.NSW has created positive opportunities within your agency? (If yes, how and can you provide any specific examples)?
    e. Do you feel that data.NSW has empowered the clients/customers of your agency? (If yes, how and can you provide any specific examples)?
    f. Do you feel that data.NSW has helped your agency solve problems? (If yes, how and can you provide any specific examples)?
    g. Has data.NSW had any other positive impacts on your agency? (If yes, how and can you provide any specific examples)?
    h. Has data.NSW had any negative impacts on your agency? (If yes, how and can you provide any specific examples)?

Final Question

7) Do you have any other comments or questions?

ICT and Digital Government

Address: McKell Building, 2-24 Rawson Place, Sydney NSW 2000

TTY: 1300 301 181

www.finance.nsw.gov.au