



The role of open data in a digitally driven economy



Contents

Introduction	03
Data must be open and liquid	03
Data is the new resource	03
The window for open data is closing soon	05
What is open data?	05
What is it worth?	05
What is happening in Australia?	06
Open data Institute opens	06
What do effective open data sources look like?	07
Open data in action	08
Architecture diagram	09
Designing open, liquid data	09
Resources	10
Contacts	10
Acknowledgments	10
Open formats	10
A-Z of open formats	10



Possums in your pocket

In the very near future, the people of NSW will be able to use the NSW Pocket Ecologist app on their mobiles to instantly discover information on the ecosystem they are in. Is the community rare or endangered? What unique plants and animals are present?

Using open data from BioNet, the Office of Environment and Heritage's (OEH) trusted source of biodiversity information for NSW, this app brings raw data to life, helping people connect to the environment around them.

NSW Pocket Ecologist is a prototype developed by the private sector through the statewide Apps4NSW initiative. It is just one example of the innovative use of biodiversity data made possible by OEH's biodiversity web services project.

However, the value of open data to the people of NSW extends beyond mobile apps. Other government agencies at state and local levels are looking to embed biodiversity data directly into their systems. This will ensure that timely decisions based on the most up-to-date and accurate biodiversity information can be made.

For example, a new system to streamline the processing of private native forestry applications at the NSW Environment Protection Authority is looking to use the web services to automatically assess the presence of threatened species in accordance with the codes of practice.

In realising the benefits of open data, OEH quickly realised that being successful was more than simply opening the door. Reducing the cost and complexity of integration for clients and ensuring that the data could be correctly and consistently interpreted were key project challenges.

These challenges were overcome by adopting standards that were already well-established with existing online communities of practice and support – OData for data access and Darwin Core for species sightings, the first data set to go live.

Introduction

As the global economy begins to operate in real time due to massive amounts of data becoming more available, the pressure is on governments and the private sector to release information. Combined with big-data analytics, this 'open data' – easily accessible, publicly funded, machine-readable information – is enabling technologies such as applications that tell commuters when the next bus or train will arrive.

Data must be open and liquid

Such data needs to be truly open and 'liquid'. It must be free, widely available and in a format that can be shared between different platforms. As with other economic resources, data must also be part of a robust supply chain that can reliably feed commercial product and service development and delivery. Seven sectors of the global economy are tipped to benefit immediately from open data: education, transport, consumer products, electricity, oil and gas, healthcare and consumer finance.¹

The calls for open data have come from the two-decade-long push to make public service information (PSI) more available to the public, which has funded the information

through taxes. Open data became a hot topic when US President Barack Obama signed the Open Government Directive on 21 January, 2009. In the five years since then, 40 countries now have government open data platforms.²

Data is the new resource

What began in response to calls for greater transparency of government spending and civic dialogue has now given way to a far more tangible need to compete in a global digital economy where data is a factor of production alongside capital.

Growing bands of entrepreneurs are now finding the data pockets and joining them into new products for consumers, and thus adding value to the economy. Historically, the process for doing this as part of a saleable solution has been cumbersome and lacked resilience to any changes in the data source. But recent technological developments have enabled data to be exposed by data custodians in an open, standards-based way consistent with the needs of modern mobile devices such as phones and tablets. The advent of these new application programming interfaces (APIs) has given

Introduction

rise to the term ‘The API Economy’ which Forrester describes as:

All commerce generated by the business of providing, consuming, integrating, and adding value to data (and thus often to products and services) via application programming interfaces that create economic value.³

This has generated an explosion in demand for data and for an accessible way to provide it at internet scale. Governments that have executed open data policies by dumping wheelbarrows full of raw data onto a portal for developers must now address the commercial context demanded by the broader digital economy, where data is the new resource. Local government must also be able to provide and consume data, given the reach, relevance and richness of the information it has to offer.

Lastly, when data is exposed by governments in the way now being operated in the API economy – ‘sockets’ that are well-documented, easy to understand and accessible to developers – it not only benefits individuals, developers and the broader economy, but also governments. With easy access to that very same data,

governments can connect between their currently siloed agencies to provide better services to citizens.

1. ‘Open data: Unlocking innovation and performance with liquid information’, McKinsey Global Institute, 2013, p2

2. *ibid*

3. http://blogs.forrester.com/mike_gilpin/13-08-22-want_to_join_the_api_economy_heres_how

Imagine if in 10 years ...

- Detailed information on energy and water use for every non-residential building in the central business district is available 24 hours a day, seven days a week. Building operators can benchmark their buildings’ energy efficiency and improve it. All the data is available because energy and water companies made data open and then worked with the council to promote annual awards for smart-energy buildings
- An electricity bill arrives at the MacMasters’ home but Mrs MacMaster knows it’s wrong. She checks online, and sees that their energy company, AusPower, has not credited the MacMasters with the power that their solar panels fed back into the grid. With the receipt number that AusPower gave her, Mrs MacMaster corrects the incorrect account and this feeds back to the company, thus improving the quality of its data and notifying the programmers of an error in their systems
- Farmer Bennett opens his laptop on the kitchen table to look at YourClimate government-sourced data on weather, crop yield and soil. All this information is free because the agencies agreed to open the data to the public. The farmer deduces that two paddocks are low in nitrogen so he must sow legumes before tomorrow’s rain. Bennett feeds this decision back to YourClimate and the data is updated instantly

The window for open data is closing soon

Open data – easily accessible, publicly funded, machine-readable information – can help citizens, and improve how governments work and deliver services. Australia is among the leaders of a global push to promote open data's social and economic benefits to the public. However, we need to move quickly to seize these opportunities while protecting privacy and intellectual property. To transform data into an information resource from which our economy can benefit, we need to make it easy, safe and useful for data custodians to publish their assets. Plus, data must be easy, reliable and valuable for developers to find, bind to and consume within their apps. This combination of making things easy but at the same time commercially robust and reliable will help spawn a new generation of data-driven innovation in our economy.

What is open data?

The UK-based Open Knowledge Foundation says “open data and content can be freely used, modified, and shared by anyone for any purpose”.⁴

The Open Data Institute Queensland (ODIQ) points to the 2014 Premier's Awards for Open Data⁴ as showing the innovation that is happening in Australia. Institute CEO, Maree Adshead, praises two of the winning apps that will enhance livability and ease access to government services.

The first is WikiQueensland, which combines open data with social media to help users make more informed decisions about where they choose to live, based on their priorities and the services they access.

The second is Pulse, which asks users to self-nominate areas of interest and then receive ‘push’ notifications on those topics. Pulse also helps government monitor public sentiment on policy, live and in real time.

- In the US, DataSF offers data grouped by category (such as ‘Economy and Community’) and presents data sets in full which are regularly updated. It also offers Rich Site Summary (RSS) so users can easily monitor new data updates. <https://data.sfgov.org>
- Kenya Open Data offers many valuable data types, including water point mapping information to improve post-construction sustainability of rural water supplies. It also offers RSS to keep users up to date. <https://opendata.go.ke>
- London Data Store presents data by type, along with simple and interactive ways to interpret it (visitors can explore a graphic representation, learn how data was sourced, read key points specific to each data set and explore or extract data itself through Excel downloads). <http://data.london.gov.uk>

What is it worth?

While open data sounds good, the technology and expertise needed to use that data effectively cost money, so the question is about return on investment. Researcher and economist Professor John Houghton, of Victoria University, says the cost-benefit ratio for open government data is 1:5. For every \$1 the government (federal, state or local) pays to make data open, \$5 is returned to the economy. The ratio for open geo-satellite data is even better, at 1:18-26.⁶

However, using conventional economic modelling to understand the value of opening up data in the way we advocate above may be misleading. The starting point for such investigation has usually involved the data's ability to enhance existing processes such as logistics and decision making. This threatens to use the past to predict the future and we suggest a more systems-based approach in predicting positive impact.

The digital disruption reconfiguring the global economy, by contrast, leverages combining various data sets with the end-user context to completely recast the business model and dynamics of whole industries. It is this step-change benefit that promises the most, but remains difficult to quantify.

Taken together, the incremental and step change impacts possible with open data point to a far more compelling case for opening up as much data as possible.

4. <http://opendefinition.org/>

5. <https://data.qld.gov.au/data-event/premiers-awards/about#winners>

6. <http://ands.org.au/resource/cost-benefit.html>

What is happening in Australia?

The three levels of government have acknowledged the value of open data and the importance of transparency in spending (see the Federal Government's open data link on the budget). But open data needs to progress beyond transparency. To work, open data needs to join up and be federated on national, state and local council levels for Australian businesses and organisations to derive real value.

Open Data Institute opens

In November 2014, sponsors from industry, research and academia funded a not-for-profit organisation, Open Data Institute Queensland, as the first Australian arm of the UK venture started in 2012 by Sir Tim Berners-Lee and Sir Nigel Shadbolt. The UK Open Data Institute aims to:

- Help people understand open data
- Overcome the challenges of publishing open data
- Make commercial gains from open data
- Use the best technologies to ensure open data is exploited in the best possible way as per style sheet.

The Open Data Institute (ODI) network, of which ODIQ is a part, encourages world-class experts to develop new ideas about open data. ODIQ will deliver tangible public value, says Maree Adshead, through working with developers and the private and public sectors to build supply chains and commercial outlets for open data. The institute is also training open data entrepreneurs, commissioning research to underpin open data technologies, and benchmarking open data initiatives.

Adshead says Australia needs to focus immediately on rapidly injecting more high-quality, high-value data sets into the open data economy. "It's a classic 'chicken or the egg' scenario – those who hold the data need inspiring stories which encourage faster, more targeted release of data," she says. "However, the type of high quality, inspirational outcomes can only be developed with good and valuable data sets."

The institute will focus its immediate efforts on the front end of the open data economy through an intensive education and training campaign within Queensland governments. This will include an ODI Certification program to elevate the quality and formatting of data being fed into the open data machine.

"We need urgent investment by the governments of Queensland to escalate this effort," Adshead says. "We will then work with business, start-ups, students, researchers and communities to educate them on what's possible and get on with solving business problems, creating new products and services, new businesses and new ways of doing things."

What do effective open data sources look like?

A data repository that is truly open has nine characteristics⁷, says Baden Appleyard, National Programme Director of the Australian Governments Open Access and Licensing Framework (AusGOAL). AusGOAL is a joint venture between all three levels of government. One of its aims is to help organisations manage risks when publishing data while simultaneously encouraging innovation.

The nine characteristics are:

- *Open encoding*
- *Open discovery*
- *Open linking*
- *Open query*
- *Open bulk supply*
- *Open identification*
- *Open presentation*
- *Open always*
- *Open now*

The most basic criteria for sharing open data are that it should be: machine-readable, searchable or indexable, and able to be easily reused within an application without requiring restrictive tools or technology.

Sir Tim Berners-Lee proposed a '5-Star' model⁸ in 2010 that provides practical guidance for organisations, but the technical landscape and the business drivers have changed considerably since the model's conception.

The broad adoption of open file formats now allows data to be accessed and used independently of a particular implementation. Whereas web pages have in the past been the primary landing and access point for data, the move to modern applications now requires access to data through well-considered APIs.

7. AusGOAL Qualities of Open Data www.ausgoal.gov.au/qualities-of-open-data

8. Five Star Open Data model <http://5stardata.info/>

We might consider the evolution of the 5-star scheme as follows:

1 Data is available via the internet (typically via WWW), with an open licence. For example, Zip files and PDFs get one star.

Benefit: Even basic access is better than none.

2 Data is machine-readable and in non-proprietary formats, for example CSV or XML.

Benefit: Data can be accessed by everyone – not just those who have particular tools or applications to read the data.

3 Data is accessible for querying and updating through open, broadly used protocols such as OData.

Benefit: OData is supported within many commercial applications and is useful, therefore, as a way of making data more meaningful and available to end-users.

4 Data is accessible through well-designed APIs able to support a wide variety of technology platforms, devices and applications.

Benefit: Commercial developers can combine different data programmatically into their apps for commercial benefit. Citizens can use consumer tools to access and visualise data.

5 Data is accessible via policy-enforced APIs enabling fine-grained control of access.

Benefit: With more control, data custodians are likely to share more of what they have.



Open data in action

AUSGRID Smart Grid, Smart City

www.smartgridsmartcity.com.au

The Smart Grid, Smart City project was a \$100 million Australian Government-funded project led by Ausgrid and supported by its partners. Smart-grid technologies such as smart sensors, smart meters and new back-end IT systems were installed to give homes more choice and control over their energy usage.

It also allowed the collection of comprehensive information relating to energy consumption patterns and smart grid efficacy – all of which was released as open data.

The open data is expected to influence decisions of governments, electricity providers, technology suppliers and consumers across Australia.

TripView

Top Paid Apps in iTunes ranking: 5

Google Play Store Top Chart ranking: 89

www.grofsoft.com/tripview.php

TripView lets users see Sydney train, bus and ferry timetable data on their mobile devices in real time. Users can accurately track the location of public transport in real time using GPS data provided by the NSW Government.

This app shows the potential for open data because it sources its GPS information from the Public Transport Information and Priority System (PTIPS). It also draws on the Sydney Coordinated Adaptive Traffic System (SCATS) to access the bus, route and timetable data of bus organisations and government authorities.

Weatherzone/Weatherzone Plus

Top Paid Apps in iTunes ranking: 8

Top Free Apps in iTunes ranking: 27

Google Play Store Top Chart ranking: 18

www.weatherzone.com.au

Weatherzone is one of the main providers of meteorological services in Australia. It has apps designed for both Android and iOS devices and services Network Ten, Australian Broadcasting Corporation and Bloomberg television.

The data is mostly an aggregation of content from various government-funded organisations, the most notable of which include the Bureau of Meteorology, Australian Broadcasting Corporation, and the World Meteorological Organisation. This is an excellent example of government data being used by private corporations to benefit the community.

Architecture diagram

Designing open, liquid data

Organisations require additional, specific capabilities to operate open data in the way we describe. Data, like other resources, observes a lifecycle requiring various stakeholders to support processing at various stages. The Private Data figure that follows outlines these.

First, the data must be extracted and ‘cleaned’. Data sources may incorporate inconsistent naming standards such as ‘car’ vs ‘automobile’, or may contain errors needing resolution before publishing.

Second, cleaned data needs to be discoverable through some kind of catalogue – particularly as the volume of available data grows. The Open Source CKAN product has been widely adopted for this purpose in Australia and internationally, and can be paired with content management tools such as Drupal or Joomla to build out a sophisticated resource for information seekers.

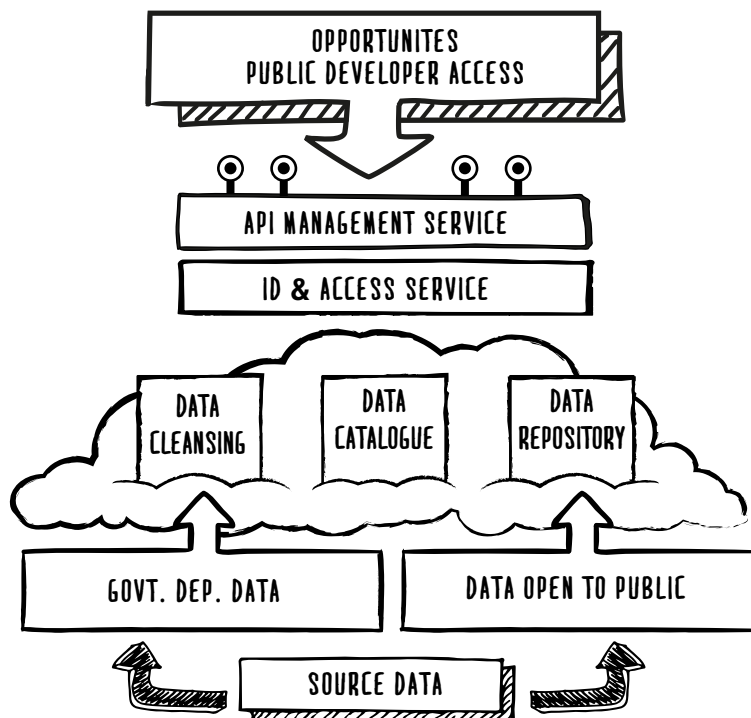
Third, discovering data is only a prelude to actually being able to bind to and consume that data using modern devices and tools. To achieve this in a way that is easy, reliable, resilient, and so can be relied on by developers considering the use of third-party data in their apps – some kind of well structured ‘front door’ is needed to stand in front of the so-called ‘raw’ data. Think of a concierge answering the door, understanding your request, and then fetching the right person for you from the midst of a very large office complex. The alternative for you is to rummage about yourself.

APIs provide machine-readable ‘front doors’ that describe how to gain authorisation to the data, and access all or part of it through copious support for open standards such as ODATA (an OASIS standard for web APIs).

Cloud-based API services achieve this and more by supporting most of the basic developer requirements such as deciding how registration is gained for the data service, how data is documented and code examples for implementing it across different devices or platforms, and importantly, ways to receive advice about any changes to the data or the service that delivers it. This level of sophistication is currently all but absent from most open data offerings.

Finally, an appropriate Identity and Access Management service ensures users of the system are uniquely identified, and any required policy is applied for those using the data. Although broad availability is a goal for open data initiatives, the integrity and availability of data is still important to both data custodians and data users and demands explicit support.

In summary then, the use of Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) support for core catalogue and content management capabilities offers a way to provide much richer access to data for the key users of these assets – developers. This approach also offers very low establishment and running costs for data providers, establishing a good road map for advancing the state of open data generally.



Resources

Contacts

AusGOAL www.ausgoal.gov.au
ODI Queensland odiqueensland.org.au
Microsoft Australia www.microsoft.com/australia
@MSFTCityNext Blog
Openness and Interoperability at Microsoft
Microsoft's Connected Government Framework
Microsoft Government Blog
ODATA BLOG

Acknowledgements

Maree Adshead, CEO, Open Data Institute Queensland
Baden Appleyard, National Programme Director,
AusGOAL (Australian Governments Open Access and
Licensing Programme)
Adam Barker, Market Strategy Lead, Microsoft Australia
Greg Stone, Chief Technology Officer, Microsoft Australia

Open formats

AusGOAL recommends the use of open formats. An open format is a specification for storing and manipulating content that is usually maintained by a standards organisation. In contrast, a proprietary format is usually maintained by a company, which then exploits the format by incorporating it into other products it sells, such as software.

Open formats are critical to the effectiveness of the open access concept. Information and data published using an open format ensures that users, regardless of their operating system or platform, can access information.

A-Z of open formats

Multimedia

JPEG 2000 – an image format standardised by ISO/IEC
Ogg – container for Vorbis, FLAC, Speex (audio formats) and Theora (a video format)
PNG – a raster image format standardised by ISO/IEC
SVG – a vector image format standardised by W3C
VRML/X3D – real-time 3D data formats standardised by ISO/IEC

Text

ASCII – a plain text file
Office Open XML – a formatted text format
OpenDocument v1.0 – a formatted text format
PDF – an open standard for documents exchange. PDF started as a proprietary standard, but was later submitted through standardisation

UTF-8 – text encoding with support for all common languages and scripts

Geospatial

KML – KML began as a proprietary standard, but was later submitted through standardisation with the Open Geospatial Consortium
WMS – Web Map Service – a protocol that allows geo-referenced map images to be served over the internet
WFS – Web Feature Service – allows requests for geographical features to be drawn across the internet

Archiving and compression

RAR – for archiving
ZIP – for both archiving and compression: the base format is in the public domain, but newer versions have some patented features

Other

CSV – comma separated values, commonly used for spreadsheets or simple databases
XML – a general-purpose markup language, standardised by W3C
HTML/XHTML – markup language for web pages
PHP – scripting and markup language for web development
RSS – syndication

Source: www.ausgoal.gov.au/open-formats