

Mashup Australia

Mashups and hackers are showing us a new way forward.

JOSE DIACONO

Recently, the federal government staged the ultimate geekfest: Mashup Australia. Developers were invited to use government datasets to build applications that would extend the use of that data.

But Mashup Australia is more than just a competition. It represents a new mindset in government circles, towards understanding that the value of data grows as it is used and re-used. Restricting information may still give its possessors power, but liberating it can also be enriching.

This article looks at the players behind the competition, the lessons learned, and asks: 'What happens next?'

A mashup is a web page or application that combines data or functionality from two or more external sources to create a new service. The term mashup implies easy, fast integration, frequently using open APIs and data sources to produce results

that were not the original reason for producing the raw source data, according to Wikipedia.

Mashup Australia was the brainchild of the Government 2.0 Taskforce. Its brief was 'to showcase how something as simple as, for example, the locations of government services or Census data can deliver benefits to the research, commercial and community sectors, and to citizens at large'.

The taskforce, which can be found at <http://gov2.net.au/>, was formed by federal finance minister Lindsay Tanner, and special minister of state Joe Ludwig, in June 2009.

The '2.0' in its title refers to Web 2.0 – applications that facilitate two-way interaction online, rather than just viewing information.

Some 15 highly experienced and visionary members were invited to staff the taskforce. They represented the private, not-for-profit and academic sectors, under the chairmanship of Dr Nicholas Gruen. The Public Service was represented by the departments of Finance and Deregulation, and Premier and Cabinet.

Members included Alan Noble, head of engineering at Google, Martin Stewart Weeks of Cisco, and Pip Marlow

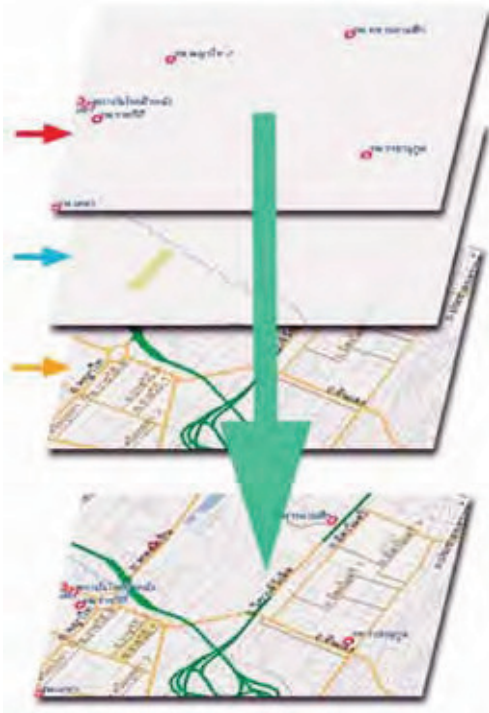
of Microsoft. Lisa Harvey of Energetica, a company that provides IT services to non-profits, was also a member.

The taskforce embarked on a roller-coaster of activities, which included roadshows, blogs, researching US and European initiatives, writing issues papers, and reviewing submissions. It finally submitted its 128-page report to the government last December, at which point it ceased to exist. The whole exercise was 'very exciting, fantastic participation and very time-consuming', according to Alan Noble.

In the spirit of open government, the operation, funding, submissions and findings of the taskforce are transparent – and well documented at the above URL. Public comments via the blog were strongly encouraged. This is the first time I have seen such lively, far-reaching and high-quality public discussion about data sharing, and the cultural change needed to support it.

The taskforce concluded quite early in the process that they needed examples, so that people could see a real life mashup. Members put out a call for data through their own contact networks. Initially 69 datasets were supplied; this later increased to 116.

National data came from the



Australian Bureau of Statistics and Bureau of Meteorology. Victoria was over-represented, with schools, hospitals, public internet locations, computer recycling ones, roadside litter collection and many more. SA had a frog atlas and crime statistics. ACT supplied playgrounds and the suburb of residence of school students.

The datasets can be viewed at <http://data.australia.gov.au>

This is a pilot website which taps into the existing infrastructure of <http://australia.gov.au>, which is run by the Australian Government Information Management Office (AGIMO).

Restricting data gives its possessors power; liberating it is enriching...

Then the taskforce enlisted the hacking community. Mashup Australia ran over a five-week period in late 2009. Mashup competitions have been held around the world, but this is believed to be the first organised by government.

The resulting 81 mashups can be seen at <http://mashupsaustralia.org/>

Lobbylens correlates data about federal government business with government suppliers and agencies, politician responsibilities, lobbyists, clients of lobbyists and the location of these entities.

Suburbmatchmaker helps people find the area of their dreams to live in.

It's bugged, mate (see <http://its-bugged-mate.apps.lpmmodules.com/>)

is designed to let people tell their local council about unserviceable items in their neighbourhood. 'Find your location on the map, then scribble on it, and leave notes about what is wrong.' This is a pilot, so all reports go into a black hole, but it is interesting to list all the things you would tell your local council if you could.

You can display different bus routes and timetables and real time traffic information on the *Darwin bus map*.

Some 56 of the 81 mashups incorporated mapping. A smattering came from traditional GIS developers, but most did not.

Several were APIs that other mashup developers could use. For instance, one application took a geocoded list of Medicare offices formatted as an .XLS address file, and converted into a geocoded ATOM web feed, suitable for Google Maps.

Three factors made the competition a success: an engaging and well-executed online infrastructure; plenty of data; and attention to the 'people side'. If any of these had been lacking, it would not have worked.

Developers need support and training (or at least some handholding). They also need motivation – cash prizes, fun, fame, or a peek inside the Google offices.

Hackfests were hosted around the country by the government, Google, OpenAustralia and Lonely Planet. They brought predominantly Open Source developers, data custodians and experts together.

Teams were formed, ideas generated, questions answered and code shared. There is an excellent video interview about the Canberra GovHack on Senator Kate Lundy's website.

Alan Noble emphasises that Mashup Australia was a trial. It was not designed as a permanent solution. Its purpose is to help those who had not previously published their data to get comfortable with the process.

Most datasets have some spatial content – for example, the address of a government office, or an xy co-ordinate of a public toilet. Updates are few and far between, but there is an RSS feed for recently added datasets and bushfire locations.

Formats are mostly CSV or KML, with the odd shape file. To encourage hackers to try some of the more complex formats, there was a prize for transforming data into a more reusable format – such as XML, JSON (JavaScript Object Notation, a lightweight data-interchange form) and KML.

Adam Kennedy and Jeffrey Candiloro were joint winners of the



Dr Nicholas Gruen chaired the Government 2.0 Taskforce.

Transformation Prize for their Geo2gov. In around a quarter of a second, Geo2gov geocodes a location – such as an address, postcode or placename – and drills that point through half a dozen spatial layers. It identifies federal, state, local and ward level locations, as well as location in the most recent census.

'It's the first time I have seen such far-reaching and high-quality public discussion...'

It returns information on your federal MP, with links to the Federal Parliamentary website, www.OpenAustralia.org/ and MyMP – to find exactly what they are up to.

'We wanted other teams to spend their time building wonderful new tools and applications, rather than trying to re-solve a difficult but solved problem,' said Candiloro.

Geo2gov might have added additional colour to many other team efforts, but the tight timescales meant that many missed the opportunity.

Government 2.0 will need more infrastructure of this type. Kennedy says: 'At the 32-hour long GovHack event, many teams spent up to two-thirds of their time overcoming problems associated with getting access to the data they wanted to use in their application.'

'Most of the applications were built largely in the last 10 hours of the event.'

'If the data was immediately accessible, in well-known and easily usable formats, we might well have doubled the productivity of the event,' says Kennedy, adding that some state governments are more helpful than others.

'Fortunately, all states are required to provide several types of geodata to

the ABS, who then release the whole compilation as Creative Commons. This allowed us to sidestep a number of problems with small or difficult state governments,' he said.

Daniela Fernandez co-developed the student prize-winning *Suburbmatchmaker* with Raul Caceres and Roberto Arias. This was one team that hit time-consuming data problems. The application uses government data to help citizens in NSW find the most suitable suburb for them to live, according to preferences about children, single people, religion, etc.

It helps people who have not published data to get comfortable with it...

It also incorporates crime statistics, but crime data in NSW is classified by LGA, not by suburb. 'Many people look up information about where they want to live based on postcode or name of suburb, but not local government area. We spent a considerable percentage of our time finding the data of suburbs and postcodes in NSW by local government area,' said Fernandez, who would have much preferred to have the same crime statistics for every state.

Suburbmatchmaker displays information and trends in text mode, on a dashboard or a map. 'A few years back, just the mapping section of the site would have been a project

in itself. Now, it is only an additional module of the Drupal Open Source Content Management System,' she said.

What is striking about the whole venture is how it builds on what exists already. The data. australia.gov.au website is built on existing infrastructure learned from *Showusabetterway* in the UK. OpenAustralia is based on a similar UK site, www.theyworkforyou.com/

The taskforce learned from mashup competitions and hackfests around the world. The hackers themselves draw heavily on the open source community.

The competition has closed, but the websites all remain as a resource, and hackers are continuing to develop their mashups. They post questions, bug reports and fixes on openaustralia-dev@googlegroups.com

The taskforce has done the spatial industry a great service. It has created a sense of urgency, excitement, drive and vision. It has looked at the big picture and the need for a lead agency, fundamental structural and cultural change in government. It acknowledges data issues such as quality, completeness and cost – but is not daunted.

Technology is racing ahead, but we must take the decision makers with us. It's one reason that mashups are important. They are also cool, useful and easy to understand. ➤

Jose Diacono is a Sydney-based freelance writer and communications consultant.

The Sizzle that Sells the Steak

An SDI may well be the key to unlocking the potential of mashups

BRAD SPENCER

Mashups are all the buzz right now, especially in that group that we 'spatially aware' people call 'the broader community'.

There are mashup competitions – and a lot of hype about how mashups can milk the benefits implicit in freely available government datasets.

But what are mashups? What makes them such a hot topic?

In the web development context, mashups have been defined in Wikipedia as 'a web page or application that combines data or functionality from two or more external sources to create a new service'.

It is part of a general movement to democratise the web...

There are many variations on this theme, but in essence it's all about allowing the web community to build and publish applications that aggregate spatial and non-spatial data in innovative, responsive and open ways. Some people refer to this as part of a general movement to democratise the web.

But hang on, you might say. We have had web mapping for years. Many organisations have used the internet to publish data for their own constituents. Some have made use of third party data as well. For instance, they have displayed their own data on top of Google Maps. Others have taken third party datasets and integrated them into their own for added richness.

So, from a GIS point of view, there is little exciting about mashups. It's business as usual. Just about any self-respecting GIS developer can build a mashup, with the right incentive and the right data.

Of course, in saying this, we ignore all the details about the delivery technologies. Quite frankly, this confuses the debate about what constitutes a real mashup.



Suburbmatchmaker: You can match your needs with your suburb.



Google

Alan Noble from Google was one of 15 people on the taskforce.



It's bugged, mate is a website where you can tell the local council things have gone pear-shaped.

There is another aspect that we should explore a little more closely. Right now, the promotion of mashup competitions (with significant prize money), has enabled non-spatial web application programmers to get hold of government spatial data, and integrate it with GoogleMaps, BingMaps, OpenStreetMap etc. The federal government has supported this. The website <http://data.australia.gov.au/> facilitates access to a range of spatial and non-spatial datasets, via file downloads.

Mashup developers download this data, transform it into whatever system they prefer, build their app and then publish it on the internet. The mashup developer also makes use of publicly available geo-services – online geocoders or gazetteers, charting and routing services, etc – that add complex functionality to a simple, easy-to-use mashup. This is a novel exercise for the non-GIS community. It is what constitutes a contemporary mashup.

People who build mashups depend on published data that is accessible...

But mashups are really about bringing data and processes together, so they can be published on other people's websites. That is, people who build mashups depend on published data resources that are accessible via web service requests.

To take a copy of government data and integrate it into an application is just a perpetuation of the existing web mapping technology, which has been with us for over 10 years.

What really sets the Web 2.0 mashup world apart is that anyone that builds a mashup has access to a vast array

of published spatial and non-spatial datasets, via open standards. But what do we mean by published? Some people assume that published means being able to physically acquire the data under Creative Commons and use it in their app.

Online publishing is different. It requires providers to expose their data or services in a way that is accessible via machine-readable web service requests, in the same way as Google and other base map suppliers. Both mechanisms can have the same apparent end – but one is profoundly different than the other.

Imagine getting GoogleMaps as a download file, or on CD. Yet many government datasets are provided in just this way – ABS, GA, PSMA, etc. It's painful to use, and it generates replications of different versions of the data.

Does the next generation of mashup providers plan to obtain the data again and again, in order to keep their app up to date?

In the case of Google, the company has its own processes for keeping the base map up to date, even though they get the data from a government (PSMA) channel. Of course, Google has a business model that justifies this.

The point is that the mashup developer can and should be anyone with minimal infrastructure. There is no requirement for a mashup jockey to actually host data.

This is why governments must urgently start to publish their data in a way that is accessible on the web, via open standards. There are software readable interfaces that allow mashup developers to access these resources without physically copying the data into their own hosted environments. They just connect to these services – as they do with GoogleMaps, etc.

The organisations that want to publish these data resources are usually the custodians; it is their responsibility to keep them current – retaining data closest to source.

I wonder how many people update their cadastral databases so often?

Just imagine how many different versions of the cadastre are currently in use. The supply model is to distribute a new version quarterly, but I wonder how many people actually update their cadastral databases so often?

The concept of mashups is not new – it is just not yet fully or systematically implemented. It represents green fields for our industry.

But what comes first: the chicken or the egg? Without a spatial data infrastructure, the potential of mashup will always be stifled. However, the current generation of mashups is showing the potential of what can be done with government data.

I believe that mashups are the sizzle that demonstrate the potential of establishing a fully operational Spatial Data Infrastructure – the steak.

So to my mind, the real challenge is not who can build the best, fastest or simplest mashup – although that is the easy thing to do right now, and it does help fuel demand.

The real challenge is to get data custodians to publish their data in an SDI-ready way. To establish this infrastructure will, of course, require vision, leadership and investment. ◆

Brad Spencer is the general manager of NuMaps, a Sydney web developer and publisher of mashups.