

From: Mark Nottingham <mnot@mnot.net>
To: Government 2.0 Taskforce <submissions@gov2.net.au>
Subject: Government 2.0 Issues Paper submission

13

August, 2009

Taskforce members,

To give my submission some context, I'll briefly summarize my experience and activities. I'm an American currently living in Melbourne, employed by Yahoo! in the United States.

Throughout my career, my focus has been on the use of standard protocols and formats to openly communicate across the Internet. For example, I was an editor of the Atom syndication format standard, and currently chair the IETF Working Group charged with revising the HTTP protocol. In the past, I spent several years contributing to the development of Web Services specifications, including SOAP, WS-Addressing, WSDL, WS-Policy, etc., as well as other technologies, such as the Platform for Privacy Preferences (P3P).

At Yahoo!, I am one of the primary authors of the company's guidelines for the Web services interfaces that are exposed both internally and to the wider world, and am regularly asked to consult on the design of such interfaces both at Yahoo! and elsewhere. As such, I've spent a great deal of time thinking about how to open up previously proprietary data.

I'm excited to see that Australia is making efforts towards opening government activities and data. However, I'm also concerned, as I've seen many pitfalls encountered when this is done in the Web 2.0 world.

Please find my answers to selected questions on the following pages, and forgive any repetition in them (hopefully you'll see this as evidence of pervading themes in my commentary). I hope that my submission can in some small way help.

Kind regards,

Mark Nottingham

Question 3:

What government information would you like to see made more freely available?

Some of the most valuable data in day-to-day life is very local. For example, it's immediately useful to know if your neighbor has made a planning application to add a new floor to their home, but currently most councils only advertise this through postal mail.

Even councils who do provide this information on their Web sites do so in a way that does not promote reuse. For example, Stonnington (my local council) does so, but only allows a search interface <<http://www.stonnington.vic.gov.au/www/html/856-planning-register-online.asp>>:

Planning Register Online

Please enter values to perform a search

Planning Application Search		Full search	Submit
Application # (eg 0123/07)	<input type="text"/>	Applicant Name	<input type="text"/>
Site Street Number	<input type="text"/>	Site Street Name	<input type="text"/>
Site Suburb	<input type="text"/>	Site Postcode	<input type="text"/>
Date Lodgement From	Day <input type="text"/> Month <input type="text"/> Year <input type="text"/>	Date Lodgement To	Day <input type="text"/> Month <input type="text"/> Year <input type="text"/>
Application Status	All <input type="text"/>	Ward	All <input type="text"/>
Keyword description search	<input type="text"/>	Proposal Type	<input type="text"/>
Meeting Date From	Day <input type="text"/> Month <input type="text"/> Year <input type="text"/>	Meeting Date To	Day <input type="text"/> Month <input type="text"/> Year <input type="text"/>
Decision Issued From	Day <input type="text"/> Month <input type="text"/> Year <input type="text"/>	Decision Issued To	Day <input type="text"/> Month <input type="text"/> Year <input type="text"/>
		List all	Reset Search

This is an excellent example of an attempt to be open that falls short of its potential.

To retrieve relevant entries, one has to agree to a terms and conditions disclaimer by ticking a box and clicking "Continue", and then manually filling out a form and submitting it. This process is time-consuming and doesn't serve the most common use case - a property owner who wants to remain informed about development near their home or investment.

If this data were truly open, it would be a simple matter for a third party to provide a service whereby a property owner was e-mailed, SMSed or IMed when a planning application was made near their home; additionally, other third-party services could "mash" this data with home sale data, planning overlays, and so forth.

However, not only is the data not technically available (i.e., it's hidden behind this search form), but it also is not legally available; the copyright terms on Stonnington's Web site <<http://www.stonnington.vic.gov.au/www/html/131-copyright.asp>> are sufficiently opaque that it's not clear if I could even offer such a service non-commercially.

This is, of course, just one example. However, I believe that a key outcome of the Taskforce should be coordination of such local authorities to open their data in ways that maximise reuse. Critically, it should be done in such a way that the data across Australia -- or even across the world -- is presented in the same fashion, to give the most opportunities for reuse.

Question 5:

What is needed to make the large volume of public sector information (a) searchable and (b) useable? And in each case, what do we do about legacy information in agencies? How might the licensing of on-line information be improved to facilitate greater re-use where appropriate?

Simply put, almost all data should be available in standard formats. It's important to distinguish between formats and interfaces (APIs) here; placing data in a well-known and easily understood format maximises its potential for reuse, whether that be search, "mash-ups" or other uses. APIs, on the other hand, require detailed, painstaking integration, and discourage reuse.

Likewise, it's necessary to distinguish between format conventions (meta-models) and concrete formats for specific tasks. XML, RDF, Atom and similar "formats" are, in reality, conventions for structuring data; they provide guidance for how to serialise data, but don't actually define the semantics of a particular application.

For example, it is not very helpful for courts to make their decisions available in "XML" or "ODF"; different jurisdictions might format the documents in completely different ways, making reuse of the data too labour-intensive to be practical. However, if an XML-based language or an ODF template is defined for those decisions, reuse is enabled.

When opening up data, it's thus important to specify more than just "XML", "ODF" or "RDF"; this will encourage the proliferation of formats, thereby raising the costs associated with consuming it.

Rather, one of the key outcomes of the Taskforce is to enable the definition of concrete formats for specific uses.

This is not to say that there isn't a place for, example, exposing RSS or Atom feeds of information that's amenable to this presentation, or simple data as CSV or JSON. However, more complex data models demand formats that have well-defined structure and semantics.

Since the Semantic Web is mentioned specifically in the Issues Paper, it's worth noting that it promises much, but has so far delivered comparatively little. While I've been extensively involved in Semantic Web work in the past (e.g., contributing to Tim Berners Lee's CWM, as well as other frameworks), I have come to believe that similar value can be delivered more easily and with less complexity by using XML or JSON as the basis of format definition; if necessary, it can be translated into RDF later (e.g., using GRDDL).

Search is a notoriously hard and expensive problem that the Australian government has no business getting into. If data is serialised into a format that can be understood and accessed by third parties, some (e.g., Google., Microsoft) will inevitably endeavor to index it.

Question 6:

How does government ensure that people, business, industry and other potential users of government information know about, and can readily find, information they may want to use, for example, the use of a consolidated directory or repository for public information?

Having a "single source of truth" is a very attractive vision, but it has so far been elusive at any scale. It has been attempted in various incarnations in the Enterprise software industry (e.g., UDDI) as well as elsewhere, and so far these attempts have met with partial success, at best.

I would put forth that the Australian government has neither the expertise or resources to make another attempt at this. Rather, by publishing data close to its source, timeliness and authority can be more reasonably assured.

Question 7:

Should governments mandate that information should be only kept and stored in open and publicly documented standards? Could such a stipulation raise costs or reduce flexibility?

In a word, yes (to both questions).

Questions 8 and 9:

What approaches should the Government use to allow information to be easily shared?

How can the initiatives and ideas of agencies be harnessed for the benefit of agencies across government? How can duplication of effort be avoided?

A key outcome of the Taskforce should be the establishment of a body that serves both to coordinate efforts, and to serve as a centre of expertise for open data.

In my experience, the myriad of decisions facing someone who wants to open their data exposes them to numerous pitfalls. Meeting those challenges requires both experience and context, and cannot be easily conveyed.

There is not a “one size fits all” solution to establishing data formats and practices (although I have seen a number of organisations attempt to decree “standards” for such interfaces, only to fail).

Rather, a centre of excellence that encourages and aids in the opening of data would promote good practice as well as reduce duplication of effort (or worse, proliferation of incompatible data format).

Question 15:

It often takes quite some time to compile and create consistent and reliable data - especially for large data sets. When is it appropriate to release limited and possible less accurate data and where is it appropriate to wait for higher quality and more extensive data? Where various principles are in some tension with each other, for instance quality and cost or timeliness, how should trade-offs be made?

Consistent and clear labeling of the reliability of the data, and possibly “streaming” a data source into separate tracks (e.g., old-but-reliable vs. recent-but-questionable).

It’s important to acknowledge that the appropriate level of reliability needs to be determined by its specific application, not its publisher. For example, someone trying to find the nearest national park with certain facilities will likely prefer timeliness over absolute reliability; however, someone entering a legal dispute with a park over its provision of facilities needs reliable (and perhaps historical) data.

Consider Google Maps; if Google waited until all of their map data were verified, it never would have seen the light of day, and many valuable applications would likewise have never come about. Google even provides directions from one point to another, even though that information is riddled with inaccuracies. Because Google manages the expectations around this data (e.g, with disclaimers), it’s still useful for some (but not all).

Question 14:

What criteria might we adopt in ensuring that agencies make data available in a reasonable time-frame? (And how might we define a ‘reasonable time-frame’?)

The most effective way to ensure this is to require that open document formats and publication of open data be integrated into the normal working process of government departments. In other words, it should not be an “extra step” to be taken after the normal work is done; instead, the default for all data should be to be online and available, with exceptions only where there are privacy or similar legal concerns.

A side effect of this approach is a higher likelihood of new, unforeseen applications of the data emerging. If the Government only exposes selected data with immediate application, it only enables those applications and their immediate derivatives. By opening up data broadly, it enables new uses which may not have been considered before.

Question 17:

What sort of public sector information should be released under what form of copyright license? When should government continue to utilise its intellectual property rights?

IPR is a substantial barrier to reuse, and it needs to be addressed correctly. For maximum reuse, the information needs to have as few restrictions as possible, although undoubtedly there are exceptions to be made.

Question 18:

When should agencies charge for access to information? Should agencies charge when they are providing value-added services? What might constitute 'value added services' (eg customisation of information)? [...]

Public information should not be charged for under any circumstances. That said, quality of service and terms of use should be developed, along with tools (e.g., rate limiting) to support them.

Beyond the most basic facilities (e.g., presenting data in HTML tables or PDFs for easy browsing), I do not believe that value-added services are a core competency of the Government, and should generally be left to third parties and public private partnerships.

This is because it's very easy to build an application that is attractive and impressive, without providing significant value. Meanwhile, there is a considerable amount of work to be done in opening the data and processes of the Government; this is where the agencies' focus should remain.

That said, if the government can identify opportunities where there is clearly demonstrated value and there is confidence that a project is achievable, it should aggressively pursue them, to illustrate the value of open government and encourage more development.

Question 19:

How can government take advantage of public private partnerships to increase access to public sector information without unduly constraining opportunities for third parties to use and reuse the information?

Partnerships should not have exclusive or preferential access to information under any conditions. As such, I suspect that PPPs will come to resemble competitive bidding for grants, rather than projects with special privileges.

Question 20:

What international activities relevant to this Taskforce should the Taskforce be considering and what needs to be done to improve cross-border use and interoperability of information?

Engaging in international efforts to define and potentially standardise formats for public data, as well as their use, is a significant (even critical) tool for furthering Australia's efforts to open government.

There is almost no value for reuse of data whose format is specific to one council, and very limited value to data whose format changes state-to-state. Likewise, while there is some value to having data format standards across Australia, it becomes that much more powerful when there is a global standard.

For example, jurisdictions in Australia are starting to make the results of restaurant health inspections available online <<http://www.foodauthority.nsw.gov.au/aboutus/offences/>>.

While this is useful information, it's not likely that a diner will check this information before they go to a particular restaurant, because they have to know about and navigate to a Web site specific to that jurisdiction.

If this data were Australia-wide, it would be much more likely that an existing service (whether Yahoo!7, Google Maps, UrbanSpoon or another) would use the data and present it in an easier-to-consume fashion (e.g., as part of an iPhone application) on a service the consumer is already using.

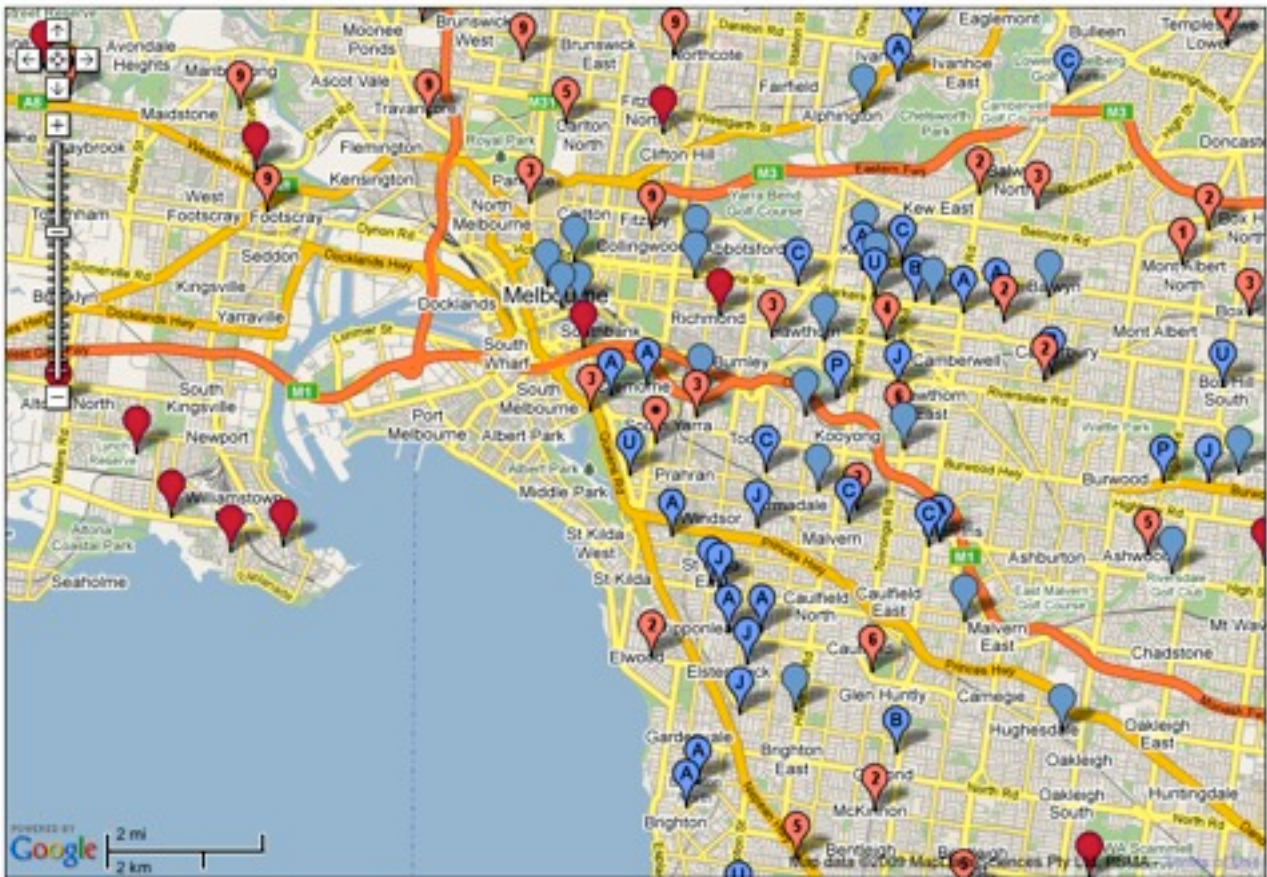
Now, consider how much more likely this reuse would be if the data were available in the same format globally; in the US, it's already common to publish this information in an ad hoc fashion. By making the data consistently available, such reuse is almost a foregone conclusion.

Practically speaking, this means that Australia should take a much more active role in standards-setting organisations such as the W3C and IETF, both to develop and encourage adoption of such formats, and to safeguard the openness of the existing Internet infrastructure.

Question 22:

Have you engaged with the Australian government via a Web 2.0 channel? Which one/s? If so, why and what was your experience? If not, why not? What can be improved?

Some time ago, I created a “mash-up” of Victorian schools: <http://www.mnot.net/blog/2007/06/30/vic_schools>.



The process was slow and cumbersome, as I had to scrape and massage the data extensively. It is difficult to update, and the data is of limited value (but still much more than provided by the original source).

If the underlying data were published at a stable location in a well-defined format that was consistent across jurisdictions, this would of course have been much easier (almost trivial), and more broadly valuable.

Question 25:

How can government make it easier for people to engage on policy and other issues and make sure the opportunities are as open and accessible as possible?

There's an important distinction to be made between opening data (i.e., allowing "read access" to what's happening in government) and opening processes ("write access").

"Read access" is comparatively easy, and should be the first step in Government 2.0.

"Write access" is significantly more complex, and needs to be approached incrementally, in comparatively small steps. Such efforts should be initially limited to gathering feedback, comments and other "user-generated data"; more serious efforts should wait until there is more experience with deploying, scaling and maintaining the "read access" data.

Question 29:

What are the barriers to fostering a culture of online innovation within government? Which of those barriers should be maintained in any Government 2.0 initiatives? Which of those barriers should be removed? How should this be achieved? What different norms can or should apply to Government 2.0 efforts?

I'm sure there are many potential barriers. However, one that may be overlooked is the relationships that government has with IT software and services vendors.

In short, many (if not most) of the vendors that government uses are predisposed towards building heavyweight, complex "Enterprise" systems. While there are appropriate uses of these technologies, their design is often at odds with simple, scalable reuse of data.

In particular, such systems are often designed around the use of specific programming languages and toolsets, and assume proximity to large Enterprise data systems (e.g., database clusters).

Adapting them to distributed reuse is difficult, because the tools are not available (or are too expensive and specialised to be used), the programming languages are too heavyweight for use by many, and it's impractical to assume low-latency, high-bandwidth connection to centralised databases.

In some cases, vendors may have an active interest in discouraging reuse and increasing complexity, because they perceive it as eroding the value they provide to the government; once the data is open, they will no longer be the gatekeeper to its value.

As such, government will need to reconsider its relationships with some vendors, either re-writing conditions of contracts or finding other incentives for making data and processes simple and easy to reuse.

It should be noted that some vendors have "Web 2.0" toolkits that promise easy migration; in almost every case, these are tools for yet more lock-in and complexity.

Question 30:

To what extent can government assist the uptake of Government 2.0 by centrally providing standard business management guidance and tools to avoid agencies having to 'reinvent the wheel' when considering their own online engagement guidelines?

In my experience a prescriptive approach does not work; the issues are too subtle and a "one size fits all" approach is not workable. This was illustrated at one of my employers where an effort to establish centralised standards for Web service interfaces failed, because the standards did not address real-world use cases, and were too restrictive.

However, significant value can be realised by investing in shared infrastructure, best practices and concentrating experience.

Question 31:

How can government engage with individuals and stakeholders to support the development of innovative policies, programs, practices and service delivery? Are there good examples of where this is happening?

The Taskforce is an excellent first step. Giving it (or an effort that eventually replaces it) a continuing mandate will be necessary to demonstrate commitment by the Australian Government, and to have measurable results.

Likewise, engaging with existing services (e.g., Flickr, Yahoo! Local, Google Maps, domain.com.au) is an excellent way to kick-start innovation, and guide determination of what data should be opened first. Similarly, there are a range of tools (contests, grants, tax incentives) that the Government can use to encourage innovation from up-and-coming or unknown parties, whether they be start-ups or teenagers in their bedrooms.

Question 32:

To what extent can we promote such an approach [feedback-driven design] in the public sector and are there any examples of emerging practice?

There may be some appropriate uses, and certainly user feedback needs to be encouraged and incorporated, but investing in such technology probably isn't appropriate in most government efforts, given that the primary objective would be to enable access to information, not influence users' decisions or monetise them.

Question 33:

How can such expertise [on existing systems and government] be governed so as not to unduly stifle innovation?

I have little experience in government, but if it is anything like private enterprise, such changes need to be undertaken by influencing key decision makers and proving the value of new processes, not by decree.

Question 34:

To what degree is the opportunity for Government agencies to participate in the Web 2.0 world inhibited, or severely compromised, by issues such as security? How might this problem be overcome, in general and by individual agencies, within current legal and policy parameters and how might these parameters be changed to assist in overcoming these problems?

It is compromised only to the degree which it is underinvested in. Web security is tricky and often expensive to get right. This is one aspect where a central authority does need to be empowered to be able to intervene (both by proactively setting mandatory policy, as well as reactively addressing issues, including shutting down insecure services).

Question 35:

What role could the proposed OIC play in encouraging the development of Government 2.0? are there practical recommendations the Taskforce might make about how the OIC might best fulfil its functions in relation to optimising the dissemination of Government information?

As I've outlined in these comments, the OIC would be one appropriate venue for a centre of excellence to encourage both good policy outcomes and appropriate technology. The critical factor is likely to be how well-resources the OIC is.