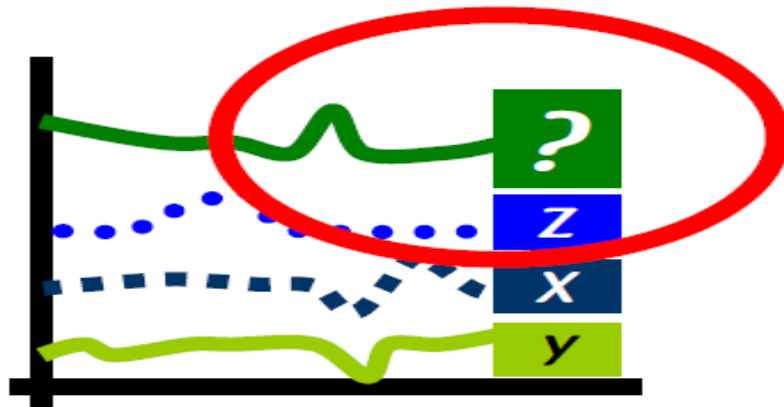


August 2009

Endless Possibilities



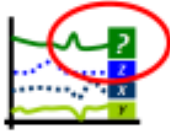
A submission to the
Government 2.0 Taskforce

Tony Cutcliffe



The Eureka Project (Aust) Pty Ltd
tc@eurekaoz.com

www.eurekaoz.com
M: 0427 535 777

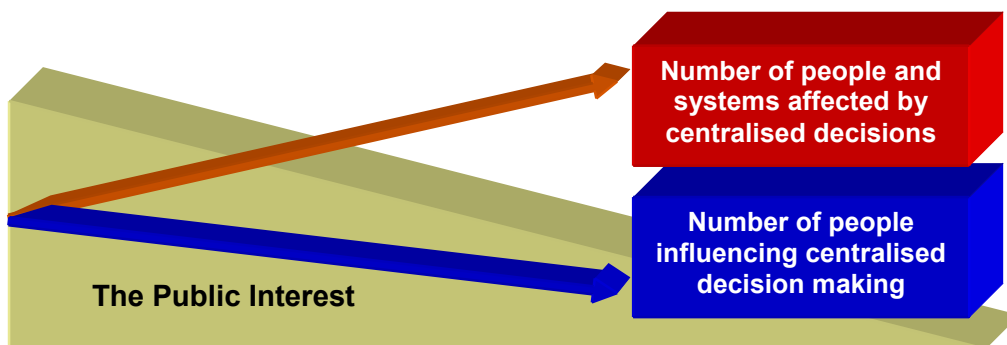


Executive Summary

In recent decades Australian communities have been re-shaped by rationalisation, privatisation and intensely centralised collection of the data that describes them. Community leadership has become overburdened and the civic conversation has become dominated by consumer aspiration. Whole societies have become passive consumers of public policy with soaring and unrealistic expectations about the quality and individualisation of government services on offer.

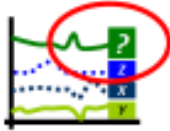
Where government has yielded to the ideology of private sector efficiencies, the corporation has thrived on oligopolistic advantages that continue to alienate the community from co-creating the shape of their own lifestyles and their country's future. All rolled together, accountability becomes lost and blurred and a palpable nihilism replaces the hope and confidence of a resilient nation.

This vicious cycle continues as the quality of decision making is deteriorated by smaller, self-referential groups making decisions for exponentially larger numbers of people with whom they have no particular relationship or empathy. Co-creation, the underlying strength of community, becomes replaced by paternalistic managerialism, and self-styled public intellectuals smother the legitimacy and value of the wisdom of crowds.



The opportunity to recreate systems and institutions is suffocated by the beneficiaries of the status quo, while the need for reinvention grows unabated and ignored. Grassroots innovation and creativity is forced to find a way around the establishments of state that were created to nurture and propagate them. The dimensions, rather than quality, of economic growth become the national obsession oblivious to the gathering and insistent demands of ecological, economic and social challenges that are simmering beneath the surface.

The *Government 2.0* taskforce has been provided with a monumentally powerful opportunity to radically improve the confidence and capacity with which Australia faces its future in a world that offers drastic constraints in availability of resources and increasingly volatile geopolitical environment. *Government 2.0* must help Australia reinvent its essential self.



Central to this proposition is the view that meaningful access to centrally collected data will provide the knowledge and thus the power to renew the outmoded system dynamics in which Australia is presently confined. Access to data is the key to the open-source environment in which achievement is built out of shared knowledge and co-creation. Collaboration and shared effort becomes the currency of progress in place of guarded intellectualism seeded in shared ignorance and power politics.

The open source environment establishes its stability by embracing change and structuring itself around instinctive relevance and the implicit knowledge of its architects. It allows communities to form around short and long term interests, problems that need solving, opportunities waiting to be identified and ideas that can grow in the humus of shared skills, experience and motivation. The collaborative culture enables problems to be designed out of systems instead of trying to design cures in.

Powered by meaningful access to data, the open source environment enables communities of interest to identify causal relationships that might never occur to conventional bureaucratic, political or commercial institutions. It creates an architecture that fine tunes interests and demand, so that policies are no longer dominated by "one size fits all" thinking, and so resources are more economically focused on delivery instead of avoiding real and imagined risks.

Most importantly this culture can avoid the necessity for major institutions to guess what their clients or constituents need or want, and for that matter, what they can do for themselves. It dispels the preoccupation and fear that comes



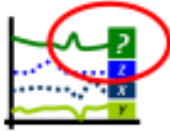
Source: java.sun.com

with zero-sum analysis that exerts its pernicious influence on policy making.

Nourished by access to real data, the open source environment also enables communities to form around particular skills and responsibilities. It provides "permission" for agents to be focused on co-creation, deployment or utilisation, and to identify and occupy complementary rather than competitive spaces. Apart from doing things better, the open source environment can identify processes and activity that have no need to continue at all. Results are valued over process and appearances.

Access to centrally collected data provides the fulcrum on which demand for scarce resources can be reliably and sustainably balanced.





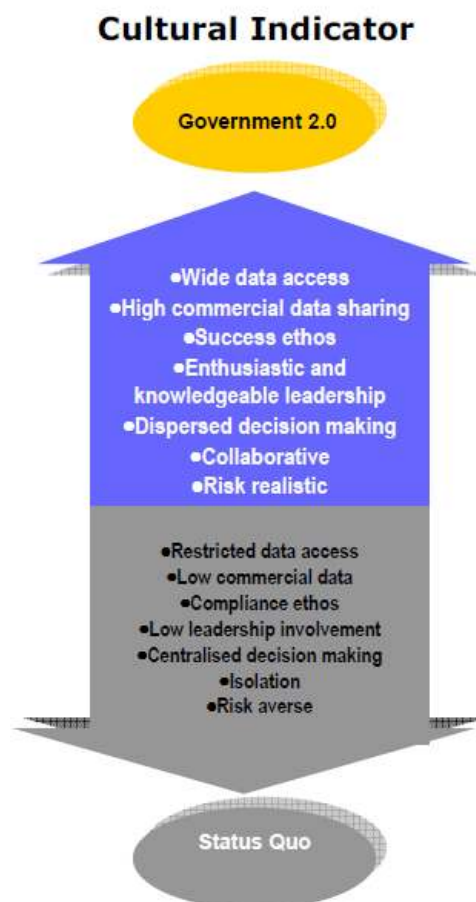
Datamining in an open source culture represents a profound evolution in policy making and will inevitably call a commensurate cultural change in the way public policy is formulated and in which regulatory concepts are developed. This cultural change should not be established on the devaluation of the way things have been done in the past, but instead be motivated by a shared and cultivated understanding of the opportunities that await. Equally, the cultural change should not diminish the value of centrally collected data, instead it should promote the value of accessing the data that is captured.

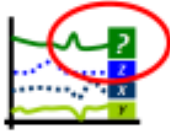
Politicians, CEOs, senior bureaucrats and the public at large must all become involved in a conversation built on a parity of worth between the contributions that each of them and their constituents has to make. Incentives should be transferred from maintaining exclusivity of knowledge to sharing data and collaborating on success. The cry of "not invented here" will be a sign of success rather than anxiety or resentment. From small beginnings and "doable" propositions will grow an energy that positively recalibrates opportunity, equity and esteem. Data access will be an investment in resilience in which all participants will develop shared and pragmatic estimations of risk and realistic expectations of themselves and others. Politicians will no longer feel the need to portray omniscience and new constructs of normality will require commercial entities to reform their perceptions of their licence to operate in the community.

Scarce funds and resources will be allocated on a radically more effective and efficient basis as the flow information takes the guesswork out of decision making.

Data access and open source applications might range from the existing widespread use of Google Earth to effective car pooling or enhanced communication styles that recognise reality instead of rhetoric. Health and education and jobs can improve as causal relationships are discovered and the quality of life, ecologies and economies will improve as people develop the power to measure what matters.

The opportunities are endless.





Some Key Points and

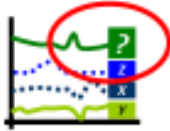
excerpts :

1. The growing array of systems, information and resource constraints is too complex for exclusively centralised decision making. Efficiency and satisfaction are linked to individualisation, intelligent defaults and opt-in and opt-out choices for consumers.
2. The extent to which decision makers are isolated from those whom their decisions affect establishes a tangent along which decisions become increasingly ineffective and resistant to popular participation. Constituents will become far more participative, and policy decisions effective, if people feel they have some say in, and control over, the policies that affect them.

Datamining should be encouraged consistently with the theory of self-determination which recognises that autonomous motivation drives productivity out of a strong sense of choice and control. When they are confronted with attempts at controlled motivation their levels of commitment, energy and participation deteriorate. This is consistent with the growing movements in Europe and the US to nurture so-called "smart communities" in a greater push for strategic relocalisation.

3. A datamining protocol must be embedded in an abiding understanding and utilisation of the open source environment. This will require deep cultural change across government as well as creative and motivational business regulation. Success will be measured by participation that goes beyond compliance.
4. Decision making aided by an open source culture will improve economic performance by designing out problems and conflicts rather than designing in cures. It will also eliminate policy guesswork and replace managerialism with acute behavioural knowledge and counter-intuitive assessments.

The Washington DC experiment with datamining and open sourcing (appsfordemocracy.org) found that applications worth US\$2.3million were developed out of a prize-pool of US\$50,000, providing a return on investment of more than 4,000%. The potential for some of the prize winning applications to reduce crime, emissions and demand for government services will produce enormous opportunities to redirect public and private funds to areas of far more productive employment.



5. Open source datamining will challenge vested interested and long established orthodoxies. Government 2.0 will need to be honest and alert to this and create strategies to overcome these obstacles.

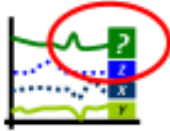
The excellent global analysis presented by Wilkinson and Pickett (*The Spirit Level* 2009 Allen Lane / Penguin) identifies the strong relationship between income disparity within communities and subsequent deterioration in levels of health, crime, education and so on. Somewhat counter-intuitively, their work indicates that soaring national GDP does nothing to alleviate this problem; in fact it can exacerbate the issue. Furthermore, they isolate a causative relationship between inequity and the British Common Law system compared with much higher levels of equity in economies that operate with the European Civil Law system.



6. The externalisation of costs and risks must be replaced with economic discussion that embraces the full economic costs and benefits of current systems and activity. This must introduce recalibration of expectations and responsibilities between the government, business and community sectors.

When gasoline cost around US\$3.00 per gallon at the bowser, the Centre estimated that this same gallon also incurred an additional cost of around \$12.00 per gallon; ie the finished cost of gasoline arrived at around \$15.00 per gallon of which 80% is subsidised by their taxes. Knowing that a tankful of gasoline for which they were charged \$50.00 was actually costing \$250.00 would provide all voters with an expectation that their government offer some better alternatives to hypermobility. Equally importantly, those systems promoters encouraging alternatives to hypermobility platforms find that they are trying to compete with systems that are supported by a hidden subsidy of 80%.

7. The open source culture can improve the quality of GDP and enlist community innovation to grow energy and resource productivity. The current burden of growing the supply of resources and services can be alleviated with new and smarter patterns of demand.
8. Datamining in an open source environment will force participants to confront deeply ingrained biases, assumptions and stereotypes. Self-knowledge is essential to developing empathy and productive relationships that will encourage creative, analysis of data and acceptance of counter-intuitive interpretations.

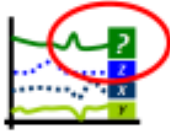


In an interesting example of such a counter-intuitive, longitudinal relationship was controversially posed by Levitt and Dubner (*Freakonomics* 2006 Penguin Books). They posed the relationship between the landmark *Roe v Wade* Supreme Court Ruling in 1973 which effectively legalised abortion in the US, and the subsequent levelling in crime rates from the early 1990s. Their analysis argues that cohorts of babies statistically destined to become criminals simply

9. The concept of "community" must be redefined to embrace any manner of associations based on geography, needs, opportunities and complementarity. The nature and variety of these associations will constantly vary and will be determined by the members of the communities themselves.
10. Governments should not try to mimic the activities of these communities of interest but should establish a position of facilitator and motivator of opportunities and responses. There must be explicit acceptance of mistakes and appreciation of learnings that will occur in the transition to an open source environment.

On 8 July 2009 the Victorian Government launched a dedicated sight on YouTube and invited questions which the Premier would answer on the basis of a poll. Two sample questions were posted by the Premier's office and one month later five original questions had been posted of which one was withdrawn. The nett result: one question per week out of a population of around 4 million Victorians. The Premier is entitled to be disappointed in the advice he received.

11. The extent to which data underlies processes, habits and systems must be consciously uncovered and evaluated with an acceptance that such knowledge is not static and will frequently be found to be outdated, incomplete or apochryphal.

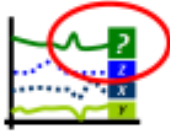


On 11 September 2001 air traffic controllers realised that commercial jetliners had been removed from the controls of their pilots. Tracking devices had been disengaged and air-control communications were replaced with silence. While logic insisted that these circumstances identified hijackings, valuable time was lost because that conclusion did not fit with the data. Historical data showed that hijackings were not conducted by people with sophisticated knowledge of aircraft operation and besides, hijackers were quick to claim their control. Despite what the live data was telling them, air traffic controllers could not believe what they were seeing (see Spencer L 2008 *Touching History* Free Press).

12. "Not invented here" should be regarded as an expression of success and celebration rather than a cause for anxiety or disaffection. The statistical validity of the wisdom of crowds must express itself in a parity of worth no matter the profile or source of a discovery, application or idea.

Proctor and Gamble realised that their complement of 9,000 researchers was a major force, yet for each one of them, there were 200 more scientists located around the world. This was regarded as a skillforce of 1.8 million professionals it could potentially utilise. P&G launched a web based "engage and connect" strategy to augment the productivity of its own researchers and also devised a patent sharing framework to incentivise qualified contributions from around the globe. The success of the initiative is reflected in the company's requirement that 50% of product and technology concepts be sourced externally by 2010.

13. The value of broad community input into system analysis and policy design will produce better results than those obtained from isolated groups of "experts". This is where the strength of open source datamining can be found.



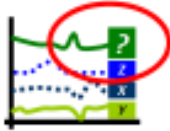
This is underpinned by the statistical theory of the "wisdom of crowds" first established by Jonathon Galt in 1912 but more recently revived by prominent researcher and writer James Surowiecki (*The Wisdom of Crowds* 2004 Random House). It establishes that a group of diverse, unconnected non-specialists will coalesce around a diagnosis or solution that is generally more effective than that of any highly qualified specialist looking at the same issue.

14. The notion of the superiority of public intellectuals in shaping policy and regulation must be challenged and exchanged for an underlying acceptance of a parity of worth.

For years, political scientists tended (or wanted) to believe that emotion-driven thinking is more characteristic of less sophisticated or less knowledgeable voters. However, the more sophisticated people are politically...the more they are able to develop complex rationalizations for dismissing data they don't want to believe. Politically knowledgeable systems also tend to be partisan, which give them the strongest reasons for distorted thinking. Work by Drew Westin (*The Political Brain* 2007 Public Affairs Press)

15. Datamining will achieve its best results when the multiplicity of human intelligences are recognised and valued. The extent to which intelligence is defined by intellectualism reduces the opportunity for innovation to be applied to data analysis and utilisation.

The value of externalised datamining is also derived from what Gardner (review his research at howardgardner.com) describes as the multiple intelligences of the human including intellectual, emotional, technical, artistic and abstract intelligences. Our present tendency to allocate "specialists" to the role of data analysis precludes the broadest possible range of causal relationships being identified.



16. Open source datamining can enable existing knowledge and experience to be leveraged in so that multiple applications of existing intellectual property can be applied. This avoids the need to reinvent the wheel and globalises the value of local knowledge.

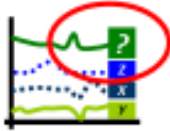
The yet2.com exchange has a network of 500 clients who are collectively estimated to hold around 40% of the world's R&D capacity who engage with around 100,000 registered participants. Clients anonymously post the technology requirement or latent capability and post the prize or contract value for successful applications provided. A car seat maker in Campellfield Victoria might choose to utilise such an exchange to identify new applications for technology it thought was useful only for carseats.

17. Knowledge is not static but the open source culture provides the opportunity to keep it current and contestable.

Nature examined 42 science entries in each of *Wikipedia* and *Encyclopaedia Britannica*. *Wikipedia* contained four inaccuracies per entry compared with *EB's* three inaccuracies per entry. The difference was of course, that *Wikipedia's* inaccuracies are soon corrected while *EB's* inaccuracies remain for the life of the product. The authors cite an MIT study which discovered that a randomly inserted obscenity is removed from *Wikipedia* in an average of 1.7 minutes.

18. Datamining in an open source environment has the potential to create platforms which provide concerted commercial yields as well as sustaining a "free agent" industry which shares in the financial success. This approach can deliver results that would not be deliverable under conventional contract arrangements.

Among the trendsetters is the extraordinarily successful Amazon.com group which has established what is arguably the most flexible and advanced on-line retail platform on the globe. Amazon created an R&D structure that has allowed 140,000 independent designers to create modules that promote and handle sales of nearly one million seller lines. Amazon provides the developers with free or at-cost access to platform modules and the developers enhance these to suit particular product and inventory applications. The developers receive a commission on every sale that is handled by their application and Amazon gets a constantly improving platform that a traditional development model could never deliver.



19. Governments can employ datamining and open source collaboration to deliver efficiencies and economies that are not available through its regular procurement and contracting arrangements. It can more effectively share risk and redirect savings for investments in services rather than administration and management.

Use of datamining in an open source environment might have allowed the government to develop its new public transport ticketing system for a tenth of the projected estimated cost of \$1.5 billion. Smart algorithmic systems used by Amazon and major US call centres might have avoided the toxic overload of the "000" telephone service during the Black Saturday bushfires.

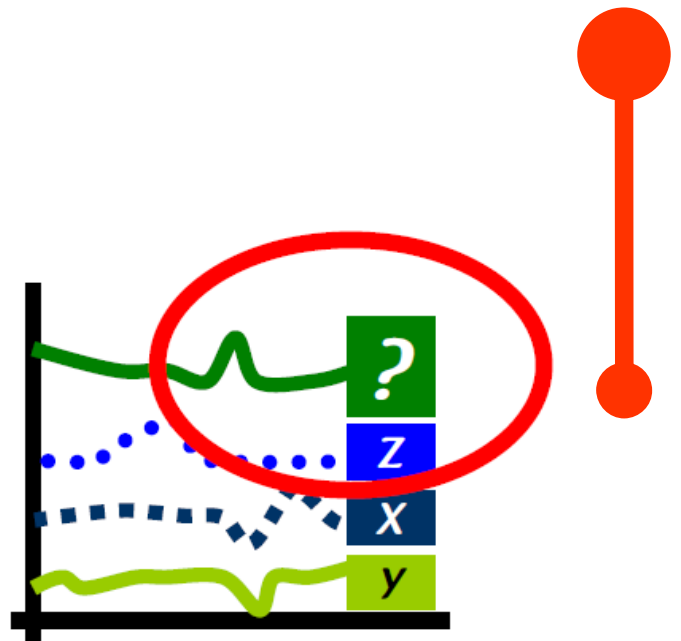
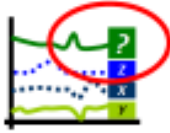
20. Sophisticated open source applications can avoid the cost of public infrastructure development while the provider's return on investment is provided from piggy-back activities.

Google Earth is one of the most powerful applications of open source software that is currently available on the internet (apart from search engines themselves). Google Earth and its overlay tools and global positioning references provided the key geospatial tools in the Victorian bushfire recovery exercise. Governments already benefit from private sector open sourcing and Government 2.0 can provide highly complementary activity.

21. The broad concept of "data" and its sheer volume create an endless array of possibilities providing that the information can be mined and applied in a collaborative, open source environment.

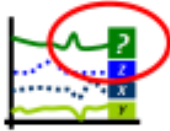
"The world is becoming interconnected. Soon, there will be two billion people on the internet – but systems and objects can now "speak" to each other, as well. Think of a trillion connected and intelligent things, and the oceans of data they will produce.

All those instrumented and interconnected things are becoming intelligent. They are being linked to powerful new backend systems that can process all that data and turn it into real insight, in real time."



Endless **Possibilities**

- Discussion•**
- Case Studies•**
- Analysis•**
- Recommendations•**
- Sources•**



Introduction

On 11 September 2001 air traffic controllers realised that commercial jetliners had been removed from the controls of their pilots. Tracking devices had been disengaged and air-control communications were replaced with silence. While logic insisted that these circumstances identified hijackings, valuable time was lost because that conclusion did not fit with the data. Historical data showed that hijackings were not conducted by people with sophisticated knowledge of aircraft operation and besides, hijackers were quick to claim their control. Despite what the live data was telling them, air traffic controllers could not believe what they were seeing (see Spencer L 2008 *Touching History* Free Press).

Such responses are typical of the human neurology. However similar lags in public policy responses can be hidden for decades under circumstances that are much slower in yielding to risk. Establishing the related cognitive and statistical theory won Daniel Kahneman (who with the late Amos Tversky) a Nobel Prize in 2002. Their decades of research showed that human decisions will generally follow a pathway of fast, automatic, short-cuts that are based on personal experiences and stereotypes. Decisions will typically only get made on the basis of rational evidence and deductive logic if the problem is effectively isolated from the alternative impulse.

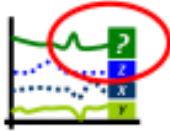
This instinctive bias is not new, but it adopts a far more significant and riskier trajectory in the present era when the data necessary for successful decision making has become so highly centralised by the nature of governments and the rapid advances in information technology. The quality of decision making has increasingly avoided effective scrutiny because communities of interest find it almost impossible to access public data and changed dynamics of social structures have undermined collectivity of determination.

Fewer and fewer people are now involved in making policy decisions which have an exponential affect on the number of people and systems they affect. Anecdotally those decisions are also being made by an increasingly narrow range of intelligences with most consideration being given political advantage instead of community and nation building.

The centralisation of decision-making reduces the opportunity for a genuine contest of ideas involved in interpreting data on which formative decisions are made. The likelihood of good ideas emerging from logical but counter-intuitive interpretation becomes lost in reversed influence of wisdom of crowds. As systems expert Peter Senge explains, this is why a room full of IQs greater than 120 have a collective IQ of 63.

The combination of these influences results in a destructive paternalism where government entities become suffocatingly protective towards their respective communities as predictable expression of knowledge exclusivity. Communities have become so remote from the data that describes them that





they are seen as needing protection from themselves. Witness the Black Saturday catastrophe in Victoria where the flow of valuable and potentially lifesaving information was inadvertently choked by highly centralised and controlling government agencies.

Work by Drew Westin (*The Political Brain* 2007 Public Affairs Press) surrounds this knowledge isolation with a further dimension of concern. His research identifies that those elite levels of decision makers deciding what is good for us are far more susceptible to woolly data than the larger masses for whom they manufacture policies:

“ For years, political scientists tended (or wanted) to believe that emotion-driven thinking is more characteristic of less sophisticated or less knowledgeable voters. However, the more sophisticated people are politically...the more they are able to develop complex rationalizations for dismissing data they don't want to believe. Politically knowledgeable systems also tend to be partisan, which give them the strongest reasons for distorted thinking. ”

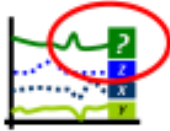
For these reasons the work of the *Government 2.0* taskforce should be greeted with enthusiasm.

The taskforce is developing a concept that needs to become the predominant influence on Australia's long-term public policy architecture. A decision to promote widespread mining of the public data inventory is a powerful invitation to reawaken community ownership, participation and leadership in designing smart responses to complex pressures confronting advanced economies.

The taskforce poses penetrating questions that must be answered in preparing recommendations designed for success. Prior to addressing those questions, this submission proposes that some preliminary considerations might be usefully considered. Perhaps also, this submission may serve as a rudimentary introduction to some of the aspects of datamining and the open source culture for members of the public seeking a doorway into the discussion.

A successful data mining protocol will radically improve the diversity of access to, and interpretation and application of, new forms of knowledge. This is in an era when restrictions of such knowledge have intimately underwritten government and commercial hegemony. It is inevitable that a redistribution of knowledge will have a material affect on the distribution of power and the privileges it affords.

Despite the most intricate rules and specifications, the data-mining protocol can only succeed if those who effectively "own" the knowledge derive a benefit from genuinely *wanting* the protocol to thrive. In the absence of such



an incentive, the protocol will be stifled by inertia generated by those who derive security from the status quo.

Creating a culture for success is arguably more important than creating the rules of the protocol. Success will emanate from flexibility, innovation and creativity nurtured and encouraged by iterative achievement. Technical compliance with the rules rather than the values of the protocol will fatally compromise the concept.

In preparing the ground for successful "supercrunching" much can be learned from the blatant circumvention of Freedom of Information (FOI) laws that presently flourishes. FOI was established as an invitation to freely access public information but its desiccated carcass now stands as a warning to inquiring minds.

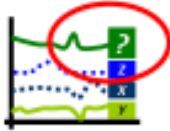
This submission approaches the cultural barrier with pragmatic optimism. It is an obstacle that is surmountable providing that it is accurately estimated from the outset. Backed by informed political will, the recommendations of the *Government 2.0* taskforce needs to bequeath Australia with inspirational and original leadership in an age of exceptional opportunity.

Cultural challenges involved in getting the data shared will be met with a corresponding challenge in getting the implications of and relationships between data believed. Datamining sits within an open-source culture and will tell us things we find hard to believe or simply don't want to believe because they will force us to change the way we think, act, earn our living, spend our money and derive our identity.

The government that spends up to \$1.5billion on a flailing transport ticketing system may find that a far superior model was available using open-sourcing for a tenth of the cost. Companies that pay tens of thousands of dollar for cash-for-chats political fundraisers may find that their access offers little benefit. Hospitals with high post-infection rates will be regarded like a restaurant with cockroaches.

The Washington DC experiment with datamining and open sourcing (appsfordemocracy.org) found that applications worth US\$2.3million were developed out of a prize-pool of US\$50,000, providing a return on investment of more than 4,000%. The potential for some of the prize winning applications to reduce crime, emissions and demand for government services will produce enormous opportunities to redirect public and private funds to areas of far more productive employment. The implications of this local initiative for national systems appear to be boundless.

The open source culture is a powerful, positive and constructive one. It is making its way around government and naturally government must get involved if it is to keep face with global developments and societal expectations. Even so, the prospect of data catalogues will inevitably be



reduced to arguments like those involving school league tables if a clever and radical campaign to change the environment is not quickly established.

Part of that environmental change involves altering the current perception and evaluation of risk. So much of Australia's lifestyle risk is currently externalised (environmental, health, crime and so) that it does not compute in the conscious mind. Open-sourcing will simply crystallise those risks but many feel that they have actually created them. Remember the Sydney Water crisis a decade ago when new testing protocols identified significant levels of *Cryptosporidium* contamination. It transpired that such contamination had always harmlessly existed, but no-one had bothered to measure it before.

Similarly, the breadth of the volumes and diversity of data is an issue with which a successful protocol will need to contend. The definition of data like the definition of influences and positions has a tendency to become entrenched rather than being recognised as an evolving and organic substance.

Consider the recent thoughtpiece published by IBM (*New Scientist* 1 August 2009) which is worth quoting at some length as it argues that data should not be treated as static and lifeless information, but as intelligence to be leveraged:

Imagine, if you can, a billion transistors for every human being. In reality, we're almost there. Sensors are being embedded everywhere: in cars, appliances, camera, roads, pipelines...even in medicine and livestock.

“

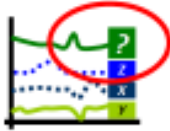
Our world is becoming interconnected. Soon, there will be two billion people on the internet – but systems and objects can now “speak” to each other, as well. Think of a trillion connected and intelligent things, and the oceans of data they will produce.

”

All those instrumented and interconnected things are becoming intelligent. They are being linked to powerful new backend systems that can process all that data and turn it into real insight, in real time.

Any person, any object, any process or service and any organisation –large or small – can become digitally aware, connected and smart.

Datamining in an open source environment will productively and positively lead to penetrating assessments of whole systems that question the most fundamental assumptions that drive policies, markets and behaviour. System dynamics will become the lingua franca of ordinary people. So-called “free agents” will frequently debunk notions of superiority claimed by self-



referential experts and public intellectuals. New forms of success will be constructed out of mutual respect and a parity of worth.

As Jay Gould remarked of public intelligence: "I am somehow less interested in the weight and convolutions of Einstein's brain than in the near certainty that people of equal talent have lived and died in cotton fields and sweatshops".

Measuring success will be instrumental to the success of the initiative. What will success look like? How will it be measured? In order to achieve a rigorous determinant of success many system costs hidden and externalised in the current lifestyle will need to be brought to account. The protocol will redefine the conventions of transparency, risk and accountability.

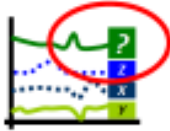
Inevitably, the consultation by the taskforce will be met with opinion and strongly held beliefs because its work is at a conceptual stage and is reaching beyond the general experiences of the community. This submission is no different and contains much of the author's personal opinion as well as fact, but it sets out to differentiate between the two.

The Government 2.0 taskforce is faced with a tremendous opportunity and this submission is intended to share some initial thoughts on the research conducted thus far by the taskforce. It is informed by the author's extensive work in knowledge networking and his subsequent confidence in the community's intrinsic resilience.

Submission Structure

This submission approaches the concept of popular datamining by noting the questions posed by the taskforce and enveloping them in layered considerations leading to the design of the protocol.





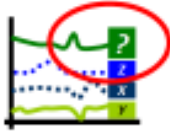
Why is popular datamining important?

Locally and within a global context Australia faces a number of competing pressures which have the opportunity to frighten the nation or to endow it with hope and optimism. In the author's view it is important to associate the concept of datamining with opportunities for positively resolving these pressures so that the concept is not dismissed as a hollow whim or elitist hobby.

The pressures are significant and might be viewed in simple terms as follows:

The Policy Pressure Cooker

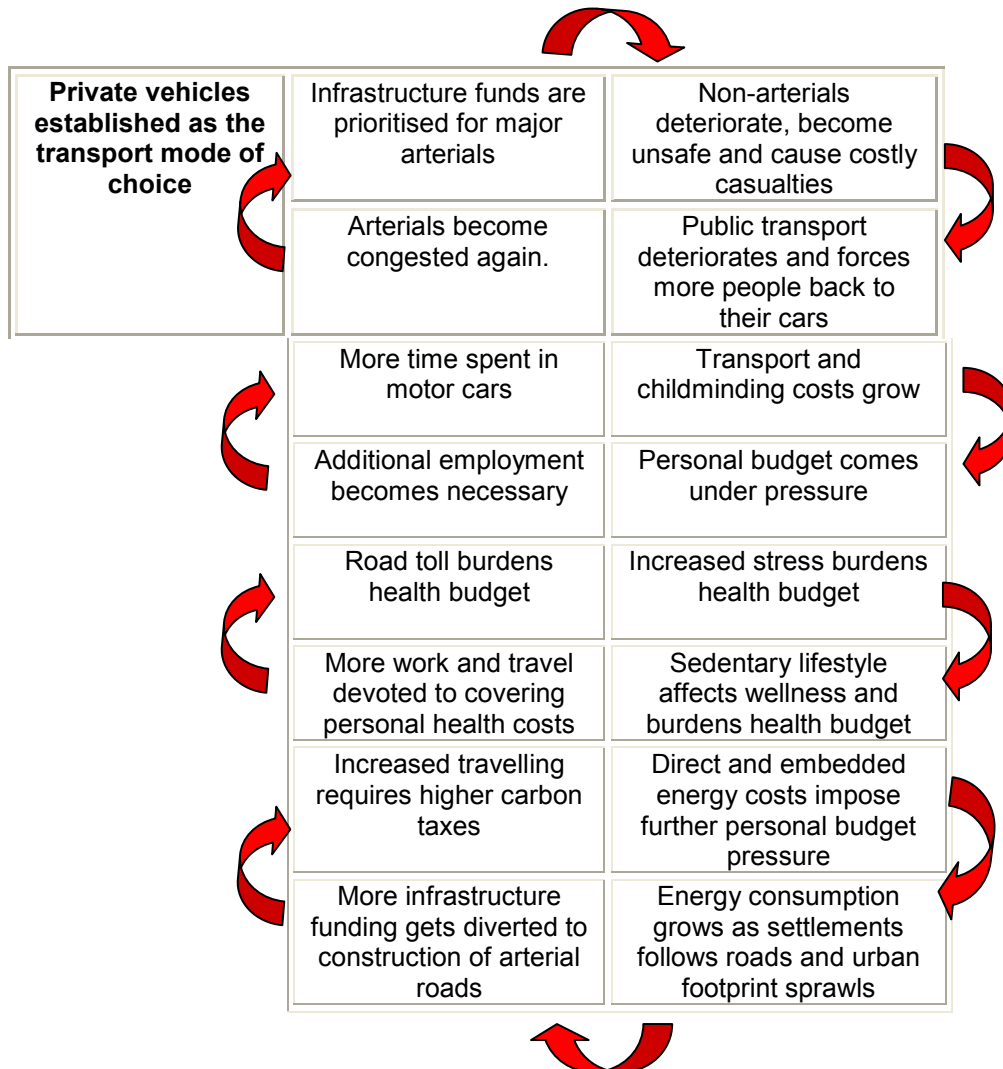




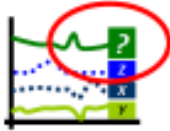
Since the mid-nineties “zero-sum government” has become the prevailing philosophical orthodoxy of the western world regardless of political stripe. It has deepened the propensity to institutionally externalise risk and accelerated the rate at which it occurs. The essential nature of the social contract between governments and their people has been based on the assumption that we can have our cake and eat it too. The discipline of substitutability will replace the threat of austerity Australia responds to the demands of a new energy economy.

Consider some of the tensions that this approach has systematically produced in the Australian context in just one area of activity alone:

The Private Motor Vehicle Conundrum



Clearly this is just a rudimentary analysis of just one of the pressures affecting policy makers and ultimately the whole economic and cultural systems on which Australians rely. Every aspect of our present lifestyle has multitudinous connections with the entire range of pressures emanating from



increasing local and global scarcity and capacity constraints involving (but not limited to):

| | | | | |
|---------------------------------|-------------------------|-----------------------------|---------------------|-----------------------------|
| Arable land | Habitable land | Water supply | Energy supplies | Materials and resources |
| Carbon emissions | Money and credit supply | Health and wellbeing levels | Growing populations | Third world development |
| Changing geo-political dynamics | Increasing expectations | Rapid change | Technology advances | Changing knowledge dynamics |



How Can Datamining Help?

The challenge for policy makers is twofold:

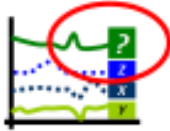
- ☑ They must identify the intersection points within this complex array of systems and devise interventions that eliminate tensions without causing real or perceived deteriorations in existing or aspirational lifestyle amenity. The interventions will clearly need to be practical, scaleable and economically, culturally and ecologically sustainable.
- ☑ They must enjoin the community's enthusiasm and capacity in identifying the intervention points and the interventions themselves so that the subsequent changes in lifestyle are seen as individually beneficial as well as satisfying systemic imperatives. The community and its administrators must be helped to recognise and distinguish between alternative pathways to lifestyle satisfaction and security.

Effective, popularised datamining provides the golden thread of knowledge that can enable these two essentials to be satisfied. There are several reasons why this is so and they require an understanding of how community dynamics have changed in the recent decades.

Community structures have been radically altered in the following respects:

Composition:

Hypermobility and advanced communication technology have introduced long-distance commuters, dormitory suburbs and transient residency. The intimacy of local knowledge and connections has been dissipated and this has been fanned by diminution of local leadership capacity arising from rationalisation and downsizing of public and commercial institutions. Increasingly, the traditional long-term employment has been replaced with radical fluidity and self-employment options. The traditional stability of community structures and membership has disappeared.



Identity:

Unconstrained by distance, the geographic basis of community has largely been replaced with more durable linear communities joined by interests, activity, occupations and consumption patterns. Communities of interest and common purpose are rapidly evolving in lieu of traditional geographic neighbourhoods.

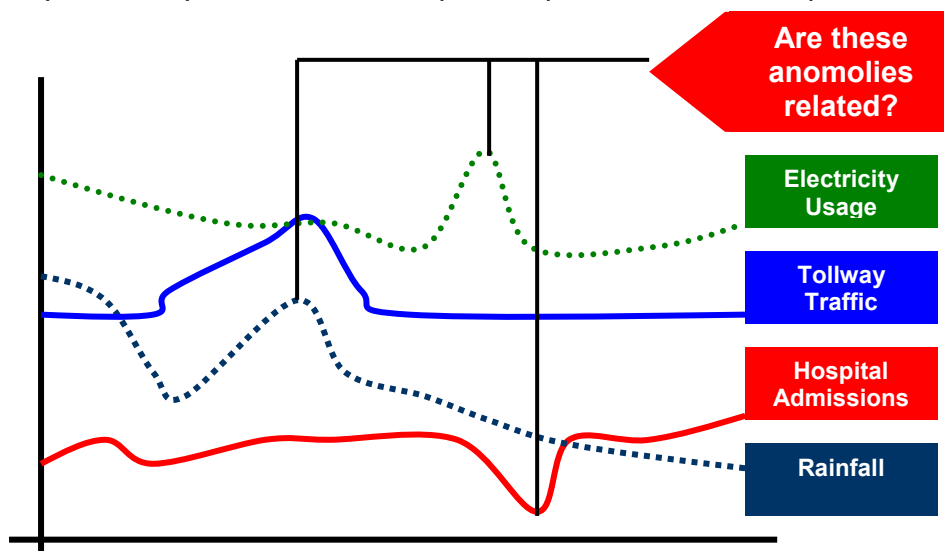
Each of these influences has produced communities that have a common characteristic: we know very little about them.

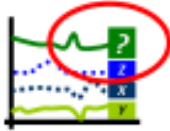
It follows therefore, that policy makers need to get to know these communities, how they operate, evolve and develop, if practical and responsive interventions points are to be devised and implemented. Datamining provides the key because it provides the opportunity for policy makers to understand these communities of interest and follow and anticipate the nature of their evolution.

Notwithstanding the growing diversity and volume of data gathered, the public policy responses which grow from it (regardless of deeper internal analysis that may have occurred) tend to be of a single dimensional nature and bereft of meaningful interconnection with other data sets. This is further entrenched by the siloistic nature of government, bureaucracy and importantly, major commercial organisations.

Macro Relationship Trending

At a macro level, datamining may help to identify causal relationships between trend patterns which are at once entirely counterintuitive but also entirely logical. By taking a helicopter view of trends it becomes possible and likely to identify intersections that present perfect intervention points:

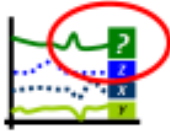




Examples of numerous significant trends found through datamining by former Clinton adviser Mark Penn (*Microtrends* 2007 Allen Lane / Penguin) have implications for behavioural predictions across multiple activities:

- ☑ US women police officers are far less likely to use, or be accused of using excessive force. The average male officer costs his jurisdiction between 2.5 and 5.5 times the female officer cost for lawsuit liability costs.
- ☑ Employees who telecommute tend to work 44.6 hours per week compared to workplace employees' 44.2 hours and 76% report job satisfaction compared to 56%.
- ☑ For every 10 minutes added to commuting time, the commuter has 10% less time for family and community activity. Longer commutes are also linked to obesity – 4/10 drivers use time spent in traffic eating and every 30 minutes committed to commuting increases the likelihood of heart disease by 3%.
- ☑ Religious groups in America that have admitted female clerics have experienced profound drops in membership.
- ☑ Around 1 in every 4 Americans looking for a life partner are doing so via internet dating sites. 17% of those introductions turn into long term relationships. Of those who marry after on-line introductions 92% say their relationship is happy and 80% say "very happy".
- ☑ In the US the average age of computer/video games players is 33 years. People under 18 years of age make up just one third of gamers, while people over 55 make up 25% and adult women at 30% are a bigger user group than boys 17 or younger at 23%.
- ☑ In the US spending on pets is now one of the top 10 retail segments. It is bigger than the confectionary, hardware and toy sectors and the top 1% of pets are believed to be receiving 40% of the purchases.
- ☑ In America today, the top 1% of pets live better than 99% of the world's population.
- ☑ American sleep studies show that men sleep less than women and African-American men sleep less with poorer sleep quality than white men and women. Sleeping just six hours per night increases the risk of obesity by 23% and just four hours per night increases the risk of obesity by 73%.

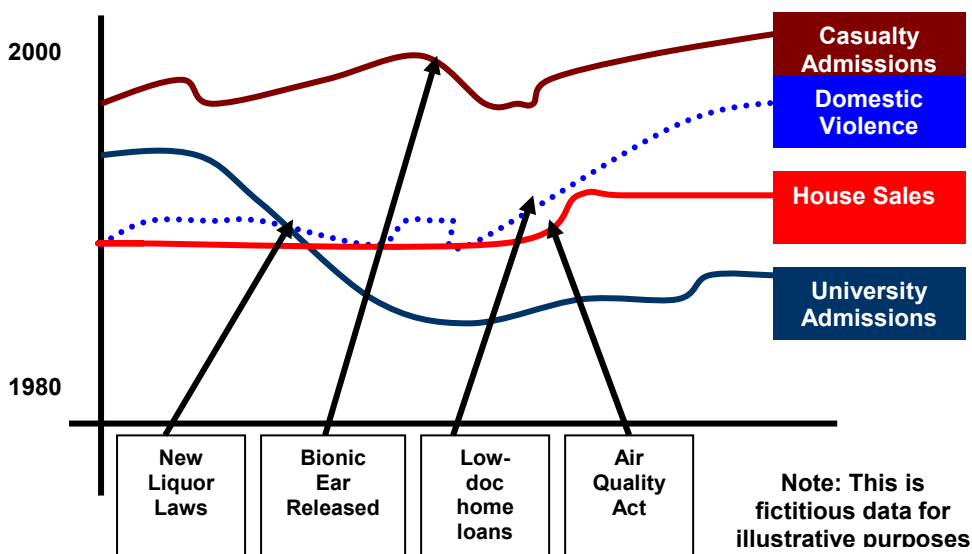




Longitudinal Event Trending

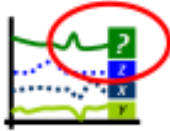
Cause and effect can take decades to manifest itself in a clear relationship yet present public discourse tends to focus inordinately on short term trends. The value of long term datamining correlated with major social, economic or legislative events can frequently identify a relationship that is usefully interpreted and reapplied:

Relationships might be highly counter-intuitive and some will take longer than others to become evident



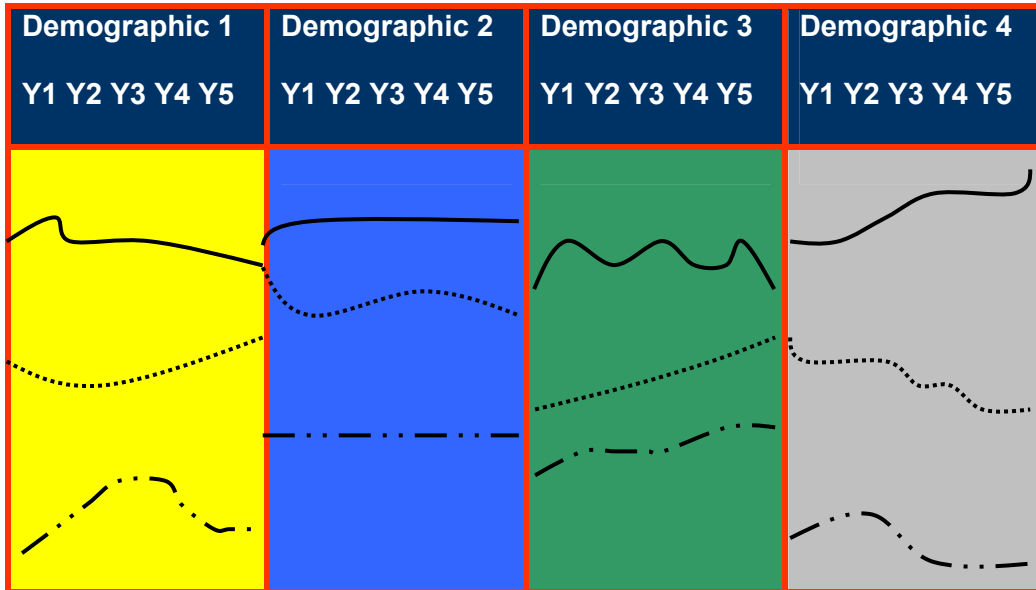
In an interesting example of such a counter-intuitive, longitudinal relationship was controversially posed by Levitt and Dubner (*Freakonomics* 2006 Penguin Books). They posed the relationship between the landmark *Roe v Wade* Supreme Court Ruling in 1973 which effectively legalised abortion in the US, and the subsequent levelling in crime rates from the early 1990s. Their analysis argues that cohorts of babies statistically destined to become criminals simply weren't born.

The excellent global analysis presented by Wilkinson and Pickett (*The Spirit Level* 2009 Allen Lane / Penguin) identifies the strong relationship between income disparity within communities and subsequent deterioration in levels of health, crime, education and so on. Somewhat counter-intuitively, their work indicates that soaring national GDP does nothing to alleviate this problem; in fact it can exacerbate the issue. Furthermore, they isolate a causative relationship between inequity and the British Common Law system compared with much higher levels of equity in economies that operate with the European Civil Law system.



Changing Values Change Habits

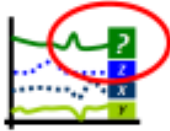
As values change among demographic groupings so do the choices they make. Datamining allows the trends of the individual groupings to be extracted from the whole and applied to the prediction of future behaviours for both policy and commercial purposes:



The work carried out by Ian Ayres (*Super Crunchers* 2007 Bantam / Random House) provides outstanding examples of specialist interpretation of data involving extraordinary amounts of data and its intimate and personalised application. This level of sophistication is evident in the existent of data specialists that warehouse and analyse data. One such firm, Teradata manages the data for 65% of the world's top retailers, 70% of airlines and 40% of banks.

Knowledge of shopping and activity habits of the different demographic groups enables major stores to finetune their inventories and pricing by day and hour. The data management chain operates from the front door of the supermarket through the entire logistics chain to growers and producers. Data crunching and the application of complex algorithms enable on-line reselling giants like Amazon to predict with powerful accuracy the sorts of books, dvds and music that individual customers may like, by cross referencing them with purchases by customers with similar tastes.





Ayres indicates that similar algorithmic devices are now in use in advanced telephone support centres of major organisations. Tripped off by the customer's account number identification, centre operators are increasingly able to answer the question before the caller actually asks it.

The value of such datamining lies in its value as an epidemiological tool that provides aggregated information that lends itself to group and individual prediction. While the majority of applications seem to emerge from the commercial arena, the implications for government services and policies are profound.

Informed Participation

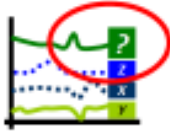
The importance of datamining is manifold in creating the architecture for Australia's future. The concept offers profound opportunities for governments, businesses, communities and individuals.

Recent decades of rationalisation and centralisation of government and commercial controls have joined with exponential development of information technology to isolate communities from self-knowledge. Urged by the market, Australians have become compliant consumers of a lifestyle that not only degrades their own wellbeing but adds to the exorbitant economic and ecological pressures confronting the nation.

The nature and dynamics of communities have also been changed extraordinarily by cheap technology, hypermobility, the flood of cheap imports and the alacrity of credit providers. From this author's experience in conducting widespread community forums, there is a sense of urgency in the desire of all forms of communities to redefine and validate their identities and take greater control over theirs and their communities' future. They demonstrate a palpable awareness that what's good for market dynamics is not necessarily good for them. Although it brings its challenges, the development is a positive one for Australia.

It should be encouraged consistently with the theory of self-determination which recognises that autonomous motivation drives productivity out of a strong sense of choice and control. When they are confronted with attempts at controlled motivation their levels of commitment, energy and participation deteriorate (see for example Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227-268). This is consistent with the growing movements in Europe and the US to nurture so-called "smart communities" in a greater push for strategic relocalisation.

The concept of relocalisation is not restricted to conventional geographic communities of towns, villages or suburbs. It can apply equally to communities of interest, the family unit or business clusters.



It is underpinned by the statistical theory of the “wisdom of crowds” first established by Jonathon Galt in 1912 but more recently revived by prominent researcher and writer James Surowiecki (*The Wisdom of Crowds* 2004 Random House). It establishes that a group of diverse, unconnected non-specialists will coalesce around a diagnosis or solution that is generally more effective than that of any highly qualified specialist looking at the same issue.

The datamining concept is not an argument against centralised collection of data; in fact it is the opposite. It urges continued and consistent aggregation of data but argues for radical new levels of access to that data so that new and meaningful interpretations and opportunities can be established.

As author Margaret Fuller succinctly proposed: “If you have knowledge, let others light their candles in it”.

The value of broader access to the data derives value from the level of intuitive, experiential and aptitudinal knowledge that “ordinary” people can bring to their analysis. Much of this knowledge, which Surowiecki refers to as “implicit knowledge”, will frequently operate at a sub-conscious or reflexive level.

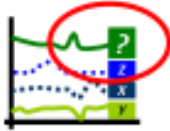
The value of externalised datamining is also derived from what Gardner (review his research at howardgardner.com) describes as the multiple intelligences of the human including intellectual, emotional, technical, artistic and abstract intelligences. Our present tendency to allocate “specialists” to the role of data analysis precludes the broadest possible range of causal relationships being identified. Dr Hugh Bradlow¹, Chief Technology Officer at Telstra once observed that some of the most valuable work at the Telstra research laboratories is conducted by anthropologists who develop alternative applications for technology developed.

This is consistent with Einstein’s view that: “The intuitive mind is a sacred gift and the rational mind is a faithful servant. We have created a society that honors the servant and has forgotten the gift”.

The creation of a datamining culture will also promote a popular demand for existing data analysis to be applied to different and important societal functions. For example, we know that data analysis by the major credit card brands offer the capacity to predict marital separation. How could this be used in a public policy setting? We know that major call centres use algorithms to predict the nature of call inquiries – why not apply this capacity to prevent overloading of the 000 emergency call centres in times of bushfire emergencies?

We know that major casinos and shopping centres use teradata to predict and encourage traffic flows and consumer behaviour. Why have we not applied this capacity to public health and sustainability campaigns?

¹ Personal conversation with the author.



The datamining concept also allows citizens the opportunity for informed participation in the public conversation. Consider the proliferation of nightclub violence in Melbourne's streets in recent years and witness the abundance of sociological musing as to why this is so. This missing piece of the jigsaw is data: anecdotal data that says the number of "at risk" licensed premises (premises serving a full range of spirits after 1.00am) has grown from around 50 five years ago to nearly 800 at present. Data relating to the approvals process shows that the effective decision making authority does not relate to the Liquor Licensing Commission or local government, but the Victorian Civil and Administrative Tribunal.

A culture of datamining can allow momentous levels of effort to be directed to creative and innovative assessment of systems rather than symptoms. It is healthy and productive for individuals, communities and organisations to properly understand the implications of system dynamics upon the way they live their lives.


Our present systems of planning, work, travel, health, law, justice and education are the legacies of a bygone era and speak to an almost boundless supply of time, resources and energy. Instead of simply addressing the supply issue, the imperative and opportunity now exists to fundamentally reshape the nature and dimensions of demand. That is too big a task for government alone and will require a collective effort, will and means to achieve it.

Datamining represents the first big step in that direction.

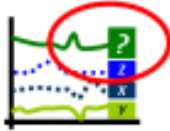
Opportunity Costs

Voters and consumers are bombarded with pressure to modify the ecological and health and wellbeing impacts of their lifestyle and a torrent of choices to match. In reality the choices don't offer much more than the likelihood of mindnumbing confusion.

The problem arises from the reality that the solutions are intended to respond to symptoms emanating from legacy systems that in many cases have their origins in the Industrial Revolution. The origins of Australia's common law system reach back into the Middle Ages.

 Datamining offers the opportunity and incentive to move beyond the symptoms and to restructure fundamental systems and the relationships between them.

The choice of diesel, hybrid or electric family cars is barely a choice at all when the systems demanding hypermobility itself are scrupulously exempted from scrutiny. Increasingly, consumer intuition begins to ask questions that go beyond the mode of transport and ponder why it is necessary at all to commute for hours every day, to use cars to get their children to school and



rely on vehicles to pursue all their shopping, entertainment and leisure pursuits.

One significant form of empowerment that enables consumers and voters to develop a far more comprehensive overview is the concept of Full Economic Cost (fEC). For the purposes of this paper, the author uses FECAB – Full Economic Cost and Benefit which recognises that some commercial and community actions offer unrewarded benefits as well as externalised costs.

Using transport as an example, the application of FECAB based on datamining provides a revealing picture that would encourage voters and consumers to demand a better range of choices to consider.

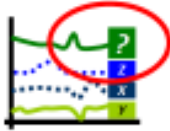
For example a recent study by the International Center for Technology Assessment considered the price of gasoline (petrol) in the US. The Center considered a variety of costs associated with the ongoing supply of petrol and diesel products for motor transport. These costs included the cost of the oil regulatory process, small particulates based lung disease, infrastructure costs, greenhouse costs and military security over international oilfields on which American oil imports rely.

At a time when gasoline costs around \$3.00 per gallon at the bowser, the Centre estimated that this same gallon also incurred an additional cost of around \$12.00 per gallon; ie the finished cost of gasoline arrived at around \$15.00 per gallon of which 80% is subsidised by their taxes. Knowing that a tankful of gasoline for which they were charged \$50.00 was actually costing \$250.00 would provide all voters with an expectation that their government offer some better alternatives to hypermobility. Equally importantly, those systems promoters encouraging alternatives to hypermobility platforms find that they are trying to compete with systems that are supported by a hidden subsidy of 80%.²



Still with transport in mind, voters might use datamining to understand the full social and economic cost of their transport commitments. As previously discussed, we know that extensive, regular commuting increases the likelihood of obesity, heart disease and stroke, and time spent commuting detracts from the person's capacity to contribute to their family and community. As far as the taxpayer is concerned, there is a reasonable expectation that their taxes should be used for national development, rather than representing an economic treadmill which the ordinary Australian powers.

² Picture: zmescience.com/breakthrough-in-biofuel-production-process



Systems evaluation based on datamining might also encourage Australians to consider the real cost benefits of the "law and order" agenda by comparing the notional savings on crime against the soaring cost of incarceration and recidivism. Householders may question the cost of having their recycling bin collected each week compared with the savings available to them by having manufacturers drastically reduce the level of packaging from the outset.

Given that the world's body of knowledge is believed to double every five years, consumers may choose to question the value for money being delivered out of the traditional "chalk and talk" methods of education and teaching.

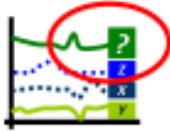
At a higher level voters may consider the value of the dollar invested in security when it is estimated that three years' US defense spending in Afghanistan and Iraq would be sufficient to provide every household in India with renewable energy. Some consumers may be pleased to find that they can reduce their weekly greenhouse gas emissions by eating vegetarian meals twice a week instead of relying on meat and its high carbon footprint.

Data interpretation may reveal to consumers that good social capital may be developed out of activities which have an entirely objective. Thomas Friedman (*Hot, Flat and Crowded* 2008 FSG Publishers) observes that US military operations in Iraq have become an effective and influential laboratory for solar energy development. Is this to reduce carbon emissions? Not really. In reality the task of moving massive diesel supplies around to all the US camps to power air-conditioners had become too risky as fuel laden convoys became increasingly attractive as targets and susceptible improvised explosive devices (IEDs). Renewable energy became important because it could save lives, but the developments will be available to the community at large.

As a matter of interest, Friedman quotes US Army energy consultant Dan Nolan who observes: "When we leave Iraq, it will be the biggest transfer of air conditioners ever known to mankind".

By examining data, consumers should be able to see what hospitals are safest from post-operative infection, while others may test the effectiveness of treatment being proposed for them. Others may be able to avoid health interventions by identifying causal relations not commonly discussed at their local clinic.

A culture of regression analysis of data can reveal an extraordinary wealth of relationships between trends, activities and outcomes. It can provide real time access to valuable information, when evidence indicates that conventional "drip-feeding" methods of distributing new medical knowledge can take up to 17 years to permeate a hospital system.



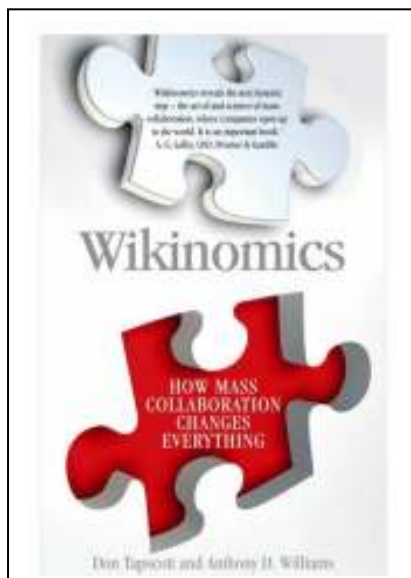
Knowledge gained from informed and purposeful datamining can be liberating in every sense of the word. Just as it can enable consumers to change their spending habits, so it can help businesses to work out what their customers might want next. Just as voters may increase their expectations of government; so governments will be able to enlist huge numbers of voters in designing and implementing new and imaginative solutions.

The way in which the protocol measures success will be a key aspect of the initiative.

Collaboration, Prosumers and Cocreation

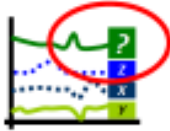
Datamining is a two-way process and relies on mutuality for success. In one direction individuals, communities and organisations will seek to interrogate public data and use it to make decisions about consumption, investments and products and services. An effective datamining culture will also create situations where the owners of data and related forms of information will proffer it up for analysis and external intellectual investment.

The new culture can create an open market for ideas, innovation and effective and logical applications of data and knowledge. This is consistent with the collaborative nature of inquiry and effort which is rapidly overtaking the traditional command and control method of devising and implementing government and commercial policy.



This is the point at which datamining converges with the concept of “wikinomics” a term popularised by Tapscott and Williams (*Wikinomics* 2006 Atlantic Books). In view of this submission the wikinomics concept must be developed in unison with datamining if the best value of either concept is to be captured. Similarly its opinion is that the text *Wikinomics* stands as the definitive guide to the wikinomic process and this section draws considerably from that text.

The authors Tapscott and Williams draw attention to the extent of collaboration being successfully encouraged and facilitated by the Internet revolution. It ushers in the era of the so-called “prosumer” where Internet users are both consumers and providers of material and content available for the cost of broadband. Rather than passively digesting information and interpretations provided by traditional institutional providers, prosumers actively shape and contest the information they trust. Command and control processes are increasingly being replaced with responsible Cocreation and the capabilities and experience of millions are harnessed in the process.



As the authors indicate, “the new art and science of wikinomics is based on four powerful new ideas: openness, peering, sharing and acting globally”.

An early showcase example that the authors use to illustrate wikinomics is provided by the teetering Canadian goldminer Goldcorp Inc. After costly geological surveys revealed the likelihood of major gold deposits existing, the company spent several years unsuccessfully trying to locate the mineable seams. After being exposed to the Linux open source platforms, the Goldcorp CEO had a radical new idea.

Instead of maintaining the usual almost paranoid secrecy of mining companies, Goldcorp decided to publish all its geological data, dating back to 1948, on the Internet and invite geologists from around the world to locate the viable seams. Launching the “Goldcorp Challenge” the company offered prizemoney of \$575,000 to entrants providing the best methods and estimates.

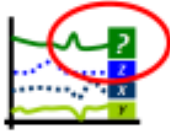
More than 1,000 “virtual prospectors” from 50 countries accepted the challenge with a diverse range of skills including geology, physics, maths, military and student participants. The contestants identified 110 targets of which 50% had not previously been considered by the company, and 80% of the targets yielded commercial gold recovery. Tapscott and Williams observe that the process transformed Goldcorp from a struggling \$100m company to a thriving organisation capitalised at more than \$9billion.

Happily, the two joint winners of the challenge were residents of Australia.

Another excellent case-study provided by Tapscott and Williams can be found in the US consumer products giant Proctor and Gamble Inc (P&G). In the late 1990s P&G realised that just 10% of the technologies being developed out of the \$1.5 billion budget were actually reaching a commercialisation stage. Furthermore the company found itself hostage to a so-called “not invented here” syndrome which excluded external input to product development.

P&G realised that their complement of 9,000 researchers was a major force, yet for each one of them, there were 200 more scientists located around the world. This was regarded as potential skillforce of 1.8 million professionals it could potentially utilise. P&G launched a web based “engage and connect” strategy to augment the productivity of its own researchers and also devised a patent sharing framework to incentivise qualified contributions from around the globe.

The success of the initiative is reflected in the company’s requirement that 50% of product and technology concepts be sourced externally by 2010.



This engage and connect concept has become hot property in the US technology and R&D sector. A further example cited in *Wikinomics* is the development of on-line technology exchanges such as yet2.com. The exchange has a network of 500 clients who are collectively estimated to hold around 40% of the world's R&D capacity who engage with around 100,000 registered participants. Clients anonymously post the technology requirement or latent capability and post the prize or contract value for successful applications provided. The participants respond within individual patent and IP protocols and both parties to the exercise exit with a satisfying transaction.

Exchanges such as yet2.com are distance neutral and provide opportunities for Australian companies searching for responses to rapidly changing markets to operate as a client or participant. An automotive seatframe maker in Campellfield Victoria might choose to reduce its exposure to the car industry by utilising such exchanges to identify new applications for technology and know how that it previously thought was useful only for carseats.

Among the trendsetters is the extraordinarily successful Amazon.com group which has established what is arguably the most flexible and advanced on-line retail platform on the globe. Amazon created an R&D structure that has allowed 140,000 independent designers to create modules that promote and handle sales of nearly one million seller lines. Amazon provides the developers with free or at-cost access to platform modules and the developers enhance these to suit particular product and inventory applications. The developers receive a commission on every sale that is handled by their application and Amazon gets a constantly improving platform that a traditional development model could never deliver.

| | Year Ended December 31, | | | | |
|--|-------------------------|------|------|------|------|
| | 2008 | 2007 | 2006 | 2005 | 2004 |

Income Statement:

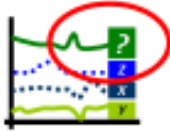
| | | | | | |
|----------------------------------|--------------------------------------|----------|----------|---------|---------|
| | (in millions, except per share data) | | | | |
| Net sales | \$19,166 | \$14,835 | \$10,711 | \$8,490 | \$6,921 |
| Income from operations | 842 | 655 | 389 | 432 | 440 |

Amazon's results speak for themselves
Source: Amazon.com 2008 Annual Report

Collaborative datamining

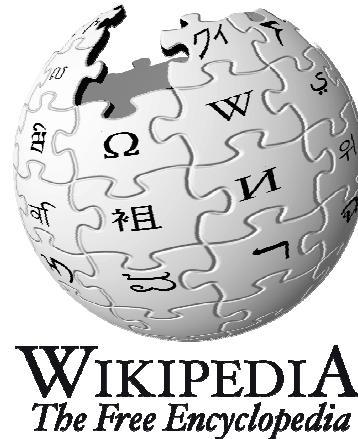
and IP exchanges have extraordinary potential in almost every field of human endeavour: health, education, logistics, knowledge management and distribution just to name a few. The dynamics of free agent activity are perhaps best illustrated with the on-line encyclopaedia *Wikipedia*. No longer must families or students but an encyclopaedia on discs (along with its singular interpretations of the facts), but they can interrogate the information and sources on *Wikipedia* and add to or correct the information available. With millions of articles in more than 200 languages, *Wikipedia* provides a dynamic and contestable knowledge forum that operates within highly





developed guidelines, supervised by moderators but largely policed by the peer pressure of contributors. Of course conventional encyclopaedia providers argued the accuracy of the *Wikipedia* content which prompted a comparative analysis by *Nature* magazine as cited by Tapscott and Williams.

Nature examined 42 science entries in each of *Wikipedia* and *Encyclopaedia Britannica*. *Wikipedia* contained four inaccuracies per entry compared with *EB's* three inaccuracies per entry. The difference was of course, that *Wikipedia's* inaccuracies are soon corrected while *EB's* inaccuracies remain for the life of the product. The authors cite an MIT study which discovered that a randomly inserted obscenity is removed from *Wikipedia* in an average of 1.7 minutes.



Tapscott and Williams identify seven key benefits to business out of harnessing peer production:

Harnessing external talent:

It is impossible for static workforces to maintain currency with the knowledge and innovations available from a global population.

Keeping up with users:

On-line providers must keep up with their users otherwise the users will "invent around them and create opportunities for competitors".

Boosting demand for complementary offerings:

Peer production creates opportunities to add value to product and service offerings and creates an affinity that increases sales of related products.

Reducing costs:

Open source collaboration delivers value beyond that available from traditional methods. IBM estimates that open source collaboration saves it \$900million pa.

Shifting the locus of competition:

The opportunity to develop non-core intellectual property that is core to a competitor can ensure that vital resources are not monopolised.

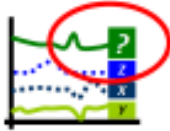
Taking the friction out of collaboration:

The standardisation of IP protocols enables efforts to be invested in R&D rather than focusing on expensive and complicated legal wrangling.

Developing social capital:

The establishment of collaborative communities provides new and innovative ways of sharing benefits, growing goodfaith and improving confidence and opportunity.

The wikinomics approach invites well founded optimism about the role which datamining can play in harvesting Australia's potential during a time of



profound change and opportunity. While collaborative datamining can have implications for government policy development, there are practical applications available for government too.

For example, consider the huge investment being made by the Victorian Government in developing a smartcard public transport ticketing system. With an original budget below \$400million the project is estimated as being three years behind schedule with estimates of the final cost currently standing at around \$1but conceivably reaching \$1.5 billion. With blue-chip organisations like IBM, P&G and Amazon combining datamining with on-line collaboration surely governments should be able to do it too?

The author of this submission acknowledges the information and analysis provided Tapscott and Williams in *Wikinomics* and used extensively in this section of the submission.

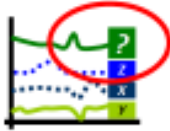
By making data and the expectations available for on-line collaboration, it is conceivable that the government would have delivered a smartcard ticketing system in half the time and at fraction of the cost. Quite apart from collaborative technical expertise, it is quite likely that some lateral thinking derived from the wisdom of crowds could have substantially reduced the complexity of the ticketing system in the first place. At a national level it seems likely that a similar approach may have helped avoid the billions of dollars bungled in defence procurement such as the aborted contract for second Seasprite helicopters.

Taxpayers may wonder if popular datamining and in-line collaboration might have delivered a longer lasting and more effective strategic planning vision than the largely failed *Melbourne 2030* plan. Equally so, global access to data might have delivered a better option to the recently let contracts for a \$3.5 billion desalination plant.

Taxpayers may also wonder if government is effective in using different interpretations of data already at its disposal. The Australian Tax Office (ATO) and the Australian Securities and Investments Commission (ASIC) will have data with excellent predictive value regarding the prospects and patterns involving a multitude of business types. How could this information be used to develop and offer advice to businesses at a major discount to the duplicated costs of state and Federal industry departments? How might this information be used to better and more finely target industry assistance as an alternative to propping up the internal combustion motor vehicle industry.

With global access to millions of minds, how might datamining and on-line collaboration design better methods of delivering TAFE and University education and eliminate the endless lags between demand and supply of necessary skills and qualifications?

Of course providing the data is available, the collaboration does not have to be on-line. The Bendigo Bank's establishment of community banking simply



matched data with faith, trust and training and numerous knock-on benefits flow out of confident and empowered communities. With pragmatic creativity and a new approach to risk management, the possibilities it seems, are endless.

A useful case study

The *Apps for Democracy* initiative supported by the District of Columbia provides a powerful insight into the benefits available from providing an open data catalogue to the assets of open-source ingenuity and implicit knowledge. It can be visited at <http://www.appsfordemocracy.org/>.

Incentivised by a \$50,000 prize pool, the scheme attracted prize-winning entries that are estimated to have avoided \$2.3million in traditional development costs. With Washington being Washington, many of the applications were devoted to improving real and perceived levels of personal safety. However the principles that were employed offer resounding indicators of the potential that exists.

The re-definition of data itself is one of the first lessons that can be gleaned from the successful exercise. Access was utilised to traditional, stored data but also utilised satellite heat maps to identify crowd levels around traditional night club and entertainment events. Parking availability was matched with live data about broken parking meters and real-time government purchasing decisions were accessed to identify purchasing patterns and locales.



The estimation of the scheme’s success by the District of Columbia is unambiguous according to Vivek Kundra, former CTO of Washington, DC and current Federal CIO: “Apps for Democracy produced more savings for the D.C. government than any other initiative.”

The underlying principle promoted by the scheme was simple and logical and profoundly elegant in its simplicity. It identified the “old” and “new” ways of doing things:

The Old Way

Few million bucks
Couple years
Crappy product

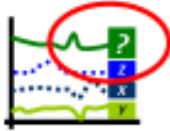
1.0 →

you know...
the way it's done now

The New Way

Open Data/Source
+
Citizen Talent
+
Fame and Fortune
=
?

2.0 →



What are the obstacles to popular datamining?

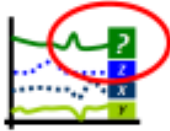
The prospects that datamining has to offer are of such a dimension that much careful attention must be paid to the potential obstacles as well benefits of popular datamining. Examining those obstacles is not an expression of negativity; it is intended as an expression of pragmatic reconnaissance. In this submission's view, the obstacles are numerous and significant but entirely surmountable. In fact there is likelihood that some of the most entrenched obstructionists could be converted into exuberant advocates.

The method of consultation employed by the taskforce itself will help set the scene for the establishment of a positive and productive culture of datamining. Many public taskforces seem primarily to attract the cognoscenti because they begin their considerations with a level of familiarity with the topic that can seem forbidding for laypeople who have an interest in making a contribution. The blog-style construction of the taskforce's website provides a useful opportunity to make contributions without involving onerous administration.

In this submission's opinion, the cultural barriers to successful datamining reinforce the need to introduce a high degree of anthropological investment in which the eventual protocol might be propagated.

Attitudes

1. There is a strong culture of secrecy involving data held by all levels of government in Australia and by government controlled utilities and corporations. While this is a longstanding culture its restrictive effect has grown as a result of the increased level of data which is centrally accumulated. While some applicants seeking information can be accused of abusing FOI processes, the entities themselves regularly appear to hide behind selective interpretations of FOI regulations, privacy regulations, commercial confidentiality and cabinet-in-confidence. Knowledge is regarded as (and actually is) the currency of authority and sharing data is seen as sharing authority and exposing policy makers to unwelcome levels of accountability. It is a serious obstacle.
2. The nature politics in Australia at present values secrecy over transparency and purpose. The extent to which short-term, poll driven policy decisions appears to motivate governments is an exact antithesis to the quality of openness and mutuality that an effective datamining protocol requires. The extent to which this culture has inculcated government and bureaucracy represents a major hurdle requiring strategic attention.



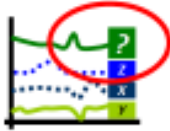
3. The effective convergence between the lobbying industry and market operators such as Private Public Partnership promoters means that political and commercial interests coincide with the effect of reinforcing each others' incentive to restrict the free flow of information.

Dataflow

1. As a rule, the annual reports required to be published by government agencies and authorities share data that is empirically and contextually compromised. Such information tends to be high level and generic providing glimpses of activity that are difficult to construe in any meaningful totality. Furthermore, the data tends to be immersed in idiom and formats peculiar to the instrumentality itself.
2. That data that is eventually released by government entities tends to be made available when its helpfulness has been diminished by age. While this may not affect some elements of longitudinal analysis its value to responsive commercial or policy decision making can be significantly degraded. Under most circumstances, knowledge is a perishable commodity and its shelf-life must be respected.
3. Government entities are not necessarily adept at measuring what matters. Entrenched processes and procedures regularly see data collected on the basis of its political potential or for purposes which have not kept pace with contemporary dynamics.
4. The growth in outsourcing, privatisation and so-called Private Public Partnerships (PPPs) means that a great deal of data previously held by government is now collected by private entities. While such information may be required for release to government under the associated contractual arrangements, a pathway to public access is rarely established.

In cases where information is released to the public the quality of its timeliness and completeness is compromised. While the taskforce is focused on mining government data, there is certain inevitability that its usefulness will require complementary data releases from the private sector.

5. There presently appears to be little or no meaningful access to live data being generated and accumulated by government systems.
6. Generally government information is provided in an aggregated basis and is not available for analysis with corresponding periodicity required for maximum utility.



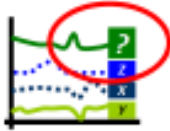
7. The level of warehoused data retrieval capacity typically available is likely to inhibit contiguity and therefore the value of analysis and conclusions available.

Media

1. Media organisations have become conditioned to dealing with Machiavellian government media units and a great deal of analysis is consequently of a short-term “gotcha” nature. Embarking on datamining journey will require cultivation of renewed goodfaith if the appropriate level of risk appetite is to be developed within government and its entities.
2. Although the datamining and on-line collaboration concepts are well advanced overseas, the required body of knowledge is in limited circulation in Australia. This is especially the case with much of Australia’s media industry and the level of preparatory knowledge that will need to be imparted should be considered as a major challenge.

Skills and Experience

1. Australia has traditionally established legal and financial careers as the stereotypical careers of choice. This has created a dearth of expertise involving *blends* of economic, mathematic, anthropological, computer and statistical skills to lead the datamining and on-line collaboration push.
2. Where such blends do exist they tend to be firmly entrenched in private sector employment where their skills and capacity are not necessarily deployed in the national interest. There will need to be a conscious effort to bolster this blended capability or else the datamining process is likely to be monopolised and manipulated by organisations with a vested interest.
3. Skills singularly based on computer skills without blended capability have a propensity to manifest themselves in highly dependant user environments. To a large extent this constraint is a generational issue and reflects management structures that do not display the intuitive adaptation that comes naturally to those who have grown up or developed with technological immersion.
4. The tools available to many prospective participants in the datamining initiative are limited in their usefulness in extracting, storing, managing and analysing data. The extent to which on-line tools are freely available (as distinct from purchasing proprietary software) to users is likely to have a major effect on the success of the datamining initiative. The principle of the “wisdom of crowds” will be an important



element of datamining success and the level of practical accessibility for participants will be a key influence on this success.

5. Knowledge is frequently confused with intelligence and therefore diminishing exclusivity of knowledge ownership can threaten positive self-perception. New perceptions of knowledge and intelligence must therefore be promoted to encourage the open data open-source culture.

What will success look like?

On 8 July 2009 Victorian Premier John Brumby launched a dedicated sight on YouTube (youtube.com/premierofvictoria) and invited questions which he would answer on the basis of a poll. Two sample questions were posted by the Premier's office and one month later five original questions had been posted of which one was withdrawn. The nett result: one question per week.

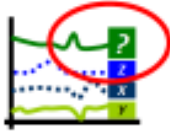
While the Premier might be applauded for tackling the on-line medium, the failure of the initiative was quite predictable. On-line communities will respond with little or no curiosity if the personality relies on conventional estimations of relevance and standing to gather a crowd. Respect gets built out of achievements that resonate with the culture and values of the on-line communities targeted. If it is clear that the intending on-line leader has simply expected to transfer the trappings and reception of the off-line community, then the initiative will disappear without trace.

In this case Premier Brumby is entitled to be disappointed in the quality of advice he received about establishing his on-line presence. More to the point, the *Government 2.0* taskforce will obviously be attuned to the culture implications of an effective datamining initiative. In the view of this submission, the datamining concept will succeed only if it looks, acts and works radically differently to the usual processes of Australian governments.

In fact there is emerging evidence that simply emulating existing social networking sites could be counter-productive. Evidence from researchers such as Oxford University's Susan Greenfield suggests that social networking can be harmful to synaptic development (for example visit news.bbc.co.uk/2/hi/programmes/newsnight/7909847.stm).

In essence, the Government 2.0 initiative needs to reintroduce resilience to Australian communities and enable them to become active participants in shaping their responsibilities, rather than simply remain passive and demanding consumers of government services (see for example Edwards C (2009) *Resilient Nation* Demos UK).

In examining what success might look like for the Government 2.0 initiative, this submission simply seeks to address the identified challenges with



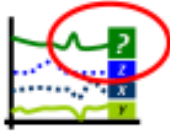
suggestions on how they might be overcome, or at least what the scene might look like in the event of a successful concept establishment. While it is not expected that any of the issues or suggestions will represent new ground for the taskforce it may be worthwhile to identify them for the benefit of public discussion.

In considering its options, the taskforce may choose to bear in mind some useful words from the systems expert Peter Senge:

“ It's common to say that trees come from seeds. But how can a tiny seed create a huge tree? Seeds do not contain the resources need to grow a tree. These must come from the medium or environment within which the tree grows. But the seed does provide something that is crucial : a place where the whole of the tree starts to form. As resources such as water and nutrients are drawn in, the seed organizes the process that generates growth. In a sense, the seed is a gateway through which the future possibility of the living tree emerges. ”

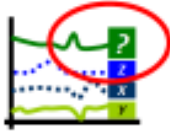


| Issue | Success Indicator |
|---------------------|--|
| Leadership | The enthusiasm and commitment to Government 2.0 by Ministers and Departmental Heads in quelling doubts and in backing the flexibility that will be required. |
| Design | Naturally Australia would do well to explore in detail the overseas experience associated with datamining and open source applications. It would be ineffective for Australia to seek to reinvent applications that have been developed and tested in countries such as the US. |
| Applications | The discussion between the taskforce and the public should clearly indicate that there are no walls around the type of applications that can be suited to Government 2.0. Applications could result in new government policies and structures; new commercial products and web-based services; new approaches to system dynamics including off-line systems; new opportunities to promote relocalisation and community resilience and new distributions of authority, responsibility and carriage between governments, business and the community. |
| Free Agents | The concept must operate on the basis of a parity of worth between participants where mutual respect is established and propagated on the basis of merit and potential. The usual should be prevented from exerting domination. |



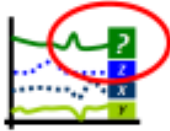
| Issue | Success Indicator |
|---------------------|---|
| Governance | A governance board should be established where principal weight is directed to behavioural, economic, collaboration and inquiry skills supported by appropriate legal, commercial and IT skills. The majority of members should be non-government employees. The board should have strong representation of overseas expertise. |
| Media | Accredited media representatives should be provided with structured tutorials on what the initiative is intended to achieve and how the open-source culture differs from conventional decision making. The media strategy must not get immersed in spin nor should it contemplate coverage being relegated to the IT "pages". |
| Incentive | Behavioural expertise should be given top priority in formulating political, bureaucratic, commercial and individual incentives to make the concept a success. This includes shaping the mode of operation of the board and its support so that they operate with full engagement in a collaborative, open source culture. |
| Skills | The initiative must specifically and overtly value the skills of collaboration, anthropology, system dynamics, behavioural economics and psychological analysis skills in lieu of the predominant trophy careers of law, investment banking and lobbying. |
| Priorities | The priorities of the protocol should be explicitly conveyed in terms of benefits to energy productivity (not energy efficiency), health and wellbeing and other externalised costs that are presently treated with end-of-pipe policies and solutions. |
| Going global | The protocol must actively seek and encourage global participation in the analysis and responses to the catalogue of data available. The protocol must take institutional steps to ensure that interest is not restricted to Australia. |
| Sanctions | In addition to incentives to release and respond to data, sanctions should be established to punish government and non-government organisations that do not release their data with the completeness and alacrity that legislation may require. The offence should be treated as seriously as |





| Issue | Success Indicator |
|-----------------------|---|
| | collusion and price fixing and the governance entity should have similar powers to the ACCC. |
| FECAB | The protocol should propagate literacy and discussion involving Full Economic Costs and Benefits associated with existing system dynamics. This is essential for creating an environment that recalibrates risk and introduce a genuine comparator between existing and proposed system arrangements. |
| Qualifications | The protocol must accept that many accomplished and productive participants will not conform with stereotypical qualification regimes. Many will have no formal qualifications, many may depart from typical age profiles and may have interests and occupations that are counter-intuitive to regular expectations. |
| Catalogue | The protocol must actively engage the on-line community and prospective participants in identifying exactly what data should be available and in what form ie live, fresh, longitudinal, latitudinal etc. |
| Benchmarks | Clear benchmarks for economic gain should be established for measuring improvements in energy productivity, health and wellbeing, human, asset and investment productivity. This will strengthen the rationale for continued development of the protocol and establish a measure of success for the protocol's design. Legislative and regulatory support should be constantly revised and improved in order to support and encourage innovation. |
| The Federation | In order to achieve the required level of cooperation from all levels of Australian government, the protocol and associated legislation should be supported by a structured co-investment program similar to the one established by Paul Keating to introduce the competition policy regime. |
| Education | The secondary and tertiary education systems are ill-equipped to support popularisation of the concepts that the protocol seeks to encourage. Appropriate incentives will need to be invested to promote cross-fertilisation of the multiple human intelligences. |





| Issue | Success Indicator |
|--------------------|--|
| Language | The language used by the taskforce and in the protocol development should be plain English, unambiguous and friendly. The initiative should not be couched in language accessible only to aficionados. Many prospectively productive participants will be scared off by language if it makes them feel inadequate or "old hat". |
| Small Steps | Peter Senge says it well: "[Seeds Are Small.] Becoming a force of nature doesn't mean that all of our aspirations must be "grand." First steps are often small, and initial visions that focus energy effectively often address immediate problems. What matters is engagement in the service of a larger purpose rather than lofty aspirations that paralyze action. Indeed, it's a dangerous trap to believe that we can pursue only "great visions." |

