PC Inquiry into Data Availability and Use

datanomics response

July 2016

Contents

- Strategic Context
- Response to PC Issues Paper
  - A "Usage-Driven" approach
    - Comment: Data vs information
    - Comment: data linking as value creation
  - Data Sharing vs Data Availability
    - Comment: a suggested 4-actor framework for data sharing
- Key data sharing priority areas
  - Public sector
  - Private sector
    - Example: SMB data sharing
    - Opportunity: Dynamic economic and industry data
  - Individuals and personal data
    - Comment: New Perspectives on the Use of Personal Data:
    - Comment: A New Lens for Strengthening Trust:
    - Opportunity: Australia Post Digital Mailbox
    - Comment: role of trusted intermediaries in personal data ecosystems
    - Example: Meeco
  - Standardisation
    - Example: Thundermaps
  - Ethics, Trust, and Privacy
    - Example: Purpose vs Interest
  - Suggested References in relation to Ethics, Privacy and Data Protection
  - About datanomics

Strategic Context

The emergence of Big Data and the subsequent proliferation of the Internet of Things (IoT) as, respectively, social and technological phenomena has generated huge transformative and disruptive forces across all industry and market sectors.

It is in this strategic context that the Productivity Commission's inquiry into Data Availability and Use should be considered.

Emerging from this is a growing recognition of some equally "heavy" issues for public and private sector organisations, and for individuals. Jeremy Rifkin[1], principle architect of the European Union's Third Industrial Revolution long-term economic sustainability plan, makes the point that in addition to the positive vision that this capture and accumulation of data presents, we will need to spend as much time dealing with what he refers to as the "chill effects" of a ubiquitously digital world. That for all the powerfully positive effects of what he has termed, the "Zero Marginal Cost Society" - the democratisation of economic life and the rise of global collaborative commons - future generations will need to wrestle with "heavy issues" about privacy, net neutrality and resilience, open and closed data systems and cybersecurity.

datanomics believe that addressing these difficult and "heavy" issues offer the greatest opportunities for sustainable social and economic value creation.

The dominant issue, however, is that of trust. More specifically, trust in terms of relationships between organisations, individuals, and things; and critically, how data and its governance, play a central role in shaping it.

Despite the rate of digital transformation of public and private sector organisations, the digital transformation of citizens and consumers they serve is accelerating an even greater rate. This, in turn, is being eclipsed by the "datafication" of "things". For the first time, humans will no longer be at the centre of the digital universe they have created. Rather, that are increasingly becoming participants in a much larger "infosphere"[2].

The emergence of this "infosphere", in which increasing amounts of business and human activities are "datafied", marks a profound shift from an industrial civilisation in which economic power is identified with ownership of the means of production, to an information one in which it economic power identified with the means of behavioural modification[3].

The enormous power of companies, like established internet giants Google, Amazon and Facebook, and newer arrivals like Uber, are their ability to "datafy" individual behaviour in order to derive economic value (i.e. advertising). This is borne out in the valuations of these organisations, which tend to reflect investor valuation of data (to drive faster revenue growth) over revenue.

[1] Jeremy Rifkin, founder of the Foundation on Economic Trends, and author - The Third Industrial Revolution
[2] Ref Luciano Floridi, Prof Philosophy of Information, Oxford Internet Institute - Fourth Revolution (2015)
[3] Shoshana Zuboff, Research article - "Big Other: surveillance capitalism and prospects of an information civilisation" (2015)

In other words, data is paramount, and data about consumers is market power for start-ups with business models and applications that allow them to amass enough to influence consumer and supplier behaviour.

datanomics observes that this transformation has occurred with relatively little consideration of its implications for individuals, society, and economies. Rather the focus has been on the successful rise of a multitude of internet giants over the past decade such as Google, Facebook, Amazon and Alibaba, and more recently "unicorns", such as Uber and Airbnb.

Equally, the dominant focus of venture capitalists and innovation programs have been on finding the next “killer app” and business model. A focus on rapid incubation and acceleration to investment and scale, have often been at the expense of startup resilience and sustainability.

datanomics believe that the strategic data risks for public and private sector organisations will arise from a number of fundamental shifts:

- Growing consumer awareness about the value of their personal data
- Growth in digital identity and data aware consumers and smaller businesses seeking new forms of value exchange
- Regulatory mindset shifts from privacy as a human right to ensuring human dignity and data ethics
- New data sharing models that enable new forms of digital engagement between businesses, public sector institutions and individuals

It is in this strategic context that datanomics believe Australia's public, private and research sectors, as well as its citizens, face major disruptive change, the economic and social impact of which will determined by the willingness of governments to pursue bold reforms in relation to the "heavy issues" associated with data and information, but also the willingness of major institutions to rise to new opportunities for data-driven collaboration and innovation.

Response to PC Issues Paper

datanomics understands that the Australian Government seeks to consider policies to increase availability and use of data to boost innovation and competition in Australia and the relative benefits and costs of each option.

In this context, datanomics believe that successful "policies to increase the availability and use of data to boost innovation and competition" need to be grounded in four principles:

> **Principle 1**: Value is not associated with data itself, but rather with the informational value realised by its usage, and the context in which it is used.
>
> *There are no high-value data sets, but rather data that contributes to high informational value use.*
>
> **Principle 2**: Data sharing is the means by which new value can be realised, both separately and collectively, by data contributors, processors, users and interests.
>
> *Policy consideration of data availability and use should occur within the wider context of data sharing in order to provide the basis for a coherent policy framework, which recognises the interests and accountabilities of all actors.*
>
> **Principle 3**: Data ethics provides the framework by which data sharing may occur in a way that respects the interests of all participants – contributors, processors, users and interests – and the legitimacy of usage.
>
> *Phenomenon such big data, advanced analytics and internet of things are challenging regulatory and legislative landscapes on issues such as personal data, legitimate purpose, agency and choice, consent and algorithmic accountability.*
>
> **Principle 4**: Data sharing policies should recognise the need for clear separation of governance between data contributors, who have rights or authority over the data, and the governance of the data technology infrastructure that supports it.
>
> *Technology does not govern data. Data technology infrastructures serve to give effect to the data governance requirements of those that hold separate or shared accountability for the data.*

In this context, datanomics believe that data sharing offers significant opportunities for creating economic and social value across all data settings, whether this is research, commercial or public.

datanomics believe that new value creation through data sharing is realised in the following ways:

- exposing new informational value from existing data sources through new usage opportunities
- creating new shared value across sources of data
- generating new informational value through derived data arising from analysis of the data
- enabling new forms of value exchange between data contributors, processors, and users
- encouraging collaborative behaviours that will realise new process and resource sharing value

datanomics considers data sharing a process, not an event. One that can only be developed and sustained through shared purposes, values and accountabilities. In doing so, data sharing enhances trust.

datanomics believe that trust in data sharing is built upon the following factors:

- **reciprocity** - the capacity for deliberate exchange that rewards data sharing and collaborative engagement
- **control** - the ability of data contributors retain control of how their data is used in the shared environment
- **individual accountability** - the willingness and capacity of data contributors to be accountable for the data they to choose to share
- **collective accountability** - the willingness and capacity of data contributors to be jointly accountable for the governance, including the matter of privacy and ethics, of shared data and its derivatives

A "Usage-Driven" approach

In addressing the core premise of the issues paper, datanomics believes that developing meaningful understandings of “the benefits and costs of options for improving the availability of and use of data” across all the sectors and stakeholders, may only be achieved in the context how the data is to be used.

Unlike physical assets, data in its raw form has no intrinsic value. Any value that may be ascribed to it can only be assessed in terms of the informational value associated with its usage. In other words, the problem it solves or questions it answers.

The value of data is determined by the informational value at the point of its consumption (or usage). Most simply put, data is much like an answer without a question. As such, information is essentially "data with a question". Without a question, the data is without meaning or value.

This means that value is determined by the question. Data that might create value in one usage context may not in another, or may result in a quite different value. **(refer comment: data vs information below)**

datanomics believe that the identification of "high-value datasets" should be considered in the context of the usage or problem domains to which they might contribute. These may be broad economic, strategic or policy domains, or specific program, project or technical in nature.

A usage-first approach also provides a sound basis for considerations about the value and nature of data linking. Decisions about data linking are inherently based on assumptions about usage, which will have impacts on its potential for informational value in other usage contexts. **(refer – comment: data linking as value creation)**

The notion of “usage domains” effectively permits framing of the data’s value in terms of economic or social value that might be expected to be derived. This also enables prospective ethical and regulatory consideration of both legitimacy of purpose, but also of use (discussed further in comments on ethics, privacy, and trust).

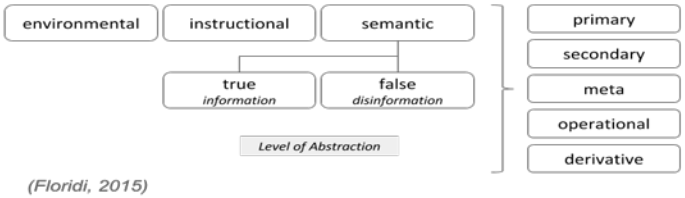
As much as the usage domain frames the nature of the information value, the origins of the data frames nature of ownership and rights associated with it. This includes issues such as rights to control, rights to benefit, availability, and accessibility. In a similar way to understanding the data usage domain, the data origination domain also enables a prospective ethical consideration about how the data might be legitimately used across usage domains.

Comment: Data vs information

While the inquiry's issues document offers a useful definition of data, the examples of data provided (e.g. personal data, big data, open data etc.) reflect a much broader usage of the term. For example, personal data and open data are not, in true definition, data, but rather they are information insofar as they have both meaning and context to some degree. Data is like an answer without a question. Information is data with a question[4]. To provide a simple illustration:

"5000 is data, but it is clearly meaningless. It could represent a numerical value, a sequence of numbers, or character, or currency. However, associated with the question, "how much is the car?" it becomes information. The question provides context. 5000 is an amount of money. The location of where the question is asked provides further context. If the question was asked in Australia, it indicates $AU as the currency. If the location is at a car dealer, the amount likely includes all additional fees (a "drive away" price) as required by Australian consumer law."

Luciano Floridi, Professor of the Ethics and Philosophy of Information at the Oxford Internet Institute offers a more thorough understanding of information. To practically illustrate this taxonomy Floridi uses the following example (ref: chapter for the Encyclopaedia of Science, Technology, and Ethics, (ESTE) edited by Carl Mitcham):



(Floridi, 2015)

"Monday morning. You turn on the ignition key of your car, but nothing happens: the engine does not even cough. Unsurprisingly, the red light of the low battery indicator is flashing. After a few more attempts, you ring the garage and explain that last night, your wife forgot to switch off the lights of the car – it is a lie, you did, but you are too ashamed to confess it – and now the battery is flat. You are told that the instruction manual of your car explains how to use jump leads to start the engine. Luckily, your neighbour has everything you need. You follow the instructions and drive to the office."

The following table applies Floridi's information taxonomy to his "flat battery" anecdote:

environmental	red light flashing
instructional	car manual
semantic (true)	battery is flat (driver's Level of Abstraction)
	12v 6-cell lead-acid (mechanic's Level of Abstraction)
	battery will cost $150-200 (economist's Level of Abstraction)
semantic (disinformation)	wife's fault *(note: misinformation is information)*
primary	no sound on turning key *(note: no information is information)*
secondary	red light flashing
meta	"battery is flat" encoded in English
operational	no red light means the battery is ok
derivative	average mileage per trip

[4] Floridi - Taxonomy of Information

This very simple example highlights two fundamental shortcomings of seeking to "identify high-value datasets":

1. the potential informational value of data depends on both the context in which exist and how it is used
2. understanding of the informational context may not be contained in a dataset

While it may seem elementary, without a sound definitional understanding of data, and recognition of the looseness of popular usage, development a coherent framework for the outcomes the inquiry is seeking to achieve will be problematic.

Comment: data linking as value creation

The concept of linking data draws its roots from records management traditions and implies a view of data as content that may be linked as one might link digital content. While linking can create value by providing a basis for associating different data sets, the process determining how the data is to be linked inherently imposes assumptions about intended its intended usage.

Approaches to data linking are varied. There are broadly two forms of data linkage:

1. **Deterministic** - which are most commonly based on using a high quality, unique identifier that is precise and stable over time (e.g. passport number or driver's licence number), or a set of rules (e.g. 100 point test)
2. **Probabilistic** - which uses some form of mathematical or computational process to link data. For example, algorithmic techniques that might be based on matching behavioural patterns (e.g. identification based on mobile call or GPS travel, health referral patterns etc)

The capacity to link of itself doesn't create value unless it is done so in a way that supports the requirements of the usage domain or problem space.

Moreover, the linking may give rise to consequential effects, with regards to privacy and re-identification. And linking may not present the linked data in a form that supports a service level required to address the usage domain. Linking may even result in distortion or degradation the informational value of the data.

Data Sharing vs Data Availability

datanomics believes that re-framing the issue of "data availability and use" as one of "data sharing" offers a richer context and more coherent basis for developing policy.

The tendency in the "data" discourse to date has been to compartmentalise data into categories that carry specific expectations, such as Open Data, Big Data, Public Data, and Personal Data etc. This has resulted in much effort trying to retrospectively ascribe meaning to them. While they offer useful terminology to describe differing phenomenon that variously share social, political, commercial and technological roots they are not based on a shared taxonomy. These categories do not discretely map upon each other, at least not so in practical terms. Their boundaries are both blurred but also contextual.

As a consequence, datanomics suggest that this conversation has been a fragmented one, with little in the way of a coherent overarching conceptual framework from which to offer a common policy foundation. However, exploration of this issue from a data sharing perspective does offer a potentially more useful approach.

As discussed earlier, datanomics believe that there are two problematic assumptions associated with a "data availability and use" oriented approach. Firstly, one that tends to the view of specific datasets as

discrete entities that can be valued as high or low; and secondly, that the implied notion of linking data is a primary approach to value creation.

The primary benefit of adopting a data sharing approach is that it allows for a more context-sensitive approach that recognises the nature of the various actors in the data sharing process, the reason(s) for data sharing, the conditions under it is being shared, and the expected outcomes. **(refer comment - a suggested 4-actor framework for data sharing)**

For example, in a data sharing context, open data is a form of sharing, where the data is shared openly for zero cost, on a "no liability" basis and with little or no usage governance. Similarly, personal data in terms of data sharing may relate to organisations sharing personal data back to individuals along with certain rights to controls and benefit from their data.

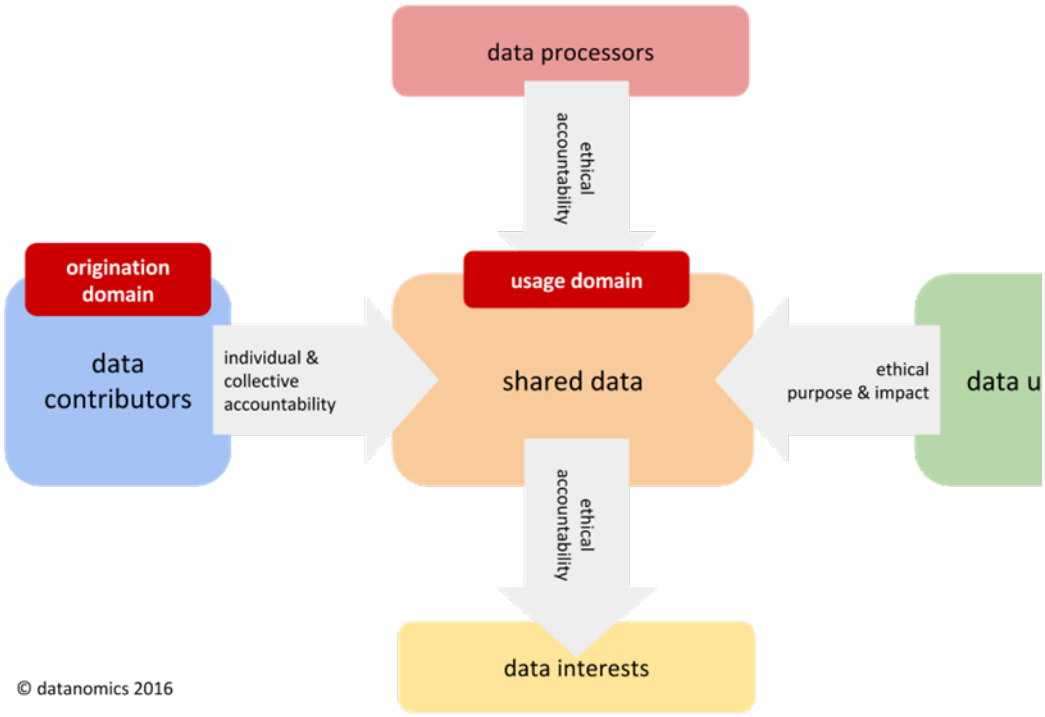
In data sharing terms, "sharing economy" and platform business models such as Uber and Airbnb, are closed "black box" data models in which suppliers and consumers are required to share data for the purposes of optimising a transaction (and in the case of Uber, also set the price), have little to no access or rights to their data, and from which the accumulated data derives increasing informational value and market power to the platform owner.

As statement of general principle datanomics holds the view that increased data sharing offers the potential for major value creation across all sectors of the Australian economy and society, (provided it is undertaken between trusted participants, whether this is within or among businesses, public sector agencies, citizens, and consumers.)

Comment: a suggested 4-actor framework for data sharing

The data sharing model comprises four actors (which may represent individuals, groups, organisations or machines):

1. **Data contributors** - have rights or authorisations to share data.
2. **Data users** - have rights or authorisations to access and use the shared data.
3. **Data processors** - have rights or authorisations to operate on the shared data.
4. **Data interests –** others that are impacted by the use of the shared data.



4-actor Data sharing framework

The nature of shared data governance is influenced by two main factors:

1. **Usage Domain** - the use for which the data is being shared. This might also be viewed as the "problem space" the data sharing is seeking to address. The Usage Domain frames how potential value is to be realised, and whose interests will be impacted.
2. **Origination Domain** - the informational context in which the data is collected. The Origination Domain frames how data may be legitimately shared and nature of governance required (ie sensitive or personal data).

Depending on the nature of data origination and usage domains, the shared data governance framework should then reflect:

- the level of individual and collective (data contributor) accountability
- ethical accountabilities of data processors and users
- collective ethical accountabilities associated with impacted parties

Whether the sharing context is that of open data, research, private, public or personal data settings there are requisite needs to understand and address these issues in an appropriate manner.

For example:

- In terms of data processing, the use of advanced ("black box") algorithms to may result in unintended and harmful consequences that may not be immediately apparent, requiring appropriate governance associated with their development and use.
- In terms of data usage, medical research demands strict data provenance requirements requiring a high degree of governance.
- In terms of data contributors, sensitive data, and personal data associated life-threatening situations require strict governance when privacy provisions need to be over-ridden.
- In terms of data interests, each of the above examples requires a level of governance capable of assessing the ethical dimensions of whose interests are being served or impacted.

datanomics

Key data sharing priority areas

Public sector

datanomics believe the following in relation to improving data availability and use in the public sector:

- that the development of usage domain oriented data sharing capabilities present all levels of government with an important opportunity to deliver improved economic and social outcomes, through more informed policy development and program delivery
- that cross agency data sharing also offers the greatest potential to understand and address increasingly more complex and systemic economic and social problems arising from the impacts of a digitally transforming economy and society
- that data sharing capabilities will be critical to supporting citizen centred digital service integration, particularly in relation to sensitive and high-risk services
- that data sharing across local and state governments sectors are particularly important given their proximity to public assets, service delivery, and citizen engagement
- that successful and sustainable data sharing will require agencies maintain separate accountability for the data they are sharing, as well as a level of collective accountability for the shared data
- that a shared governance approach is required to address trust, ethical, privacy and data protection issues associated with the usage domain

datanomics believe that issues such as data linking and standardisation should best be considered in the context of the usage domain being served. It is important to recognise that data linking effectively imposes usage assumptions about the data that may have implications based on the usage context. Similarly, approaches to data collection and standardisation should also best be considered in the context of expected usage.

datanomics believes that a "usage-driven" approach to data sharing will offer local and state government agencies opportunities to more readily address fundamental policy and program priorities by aligning it data sharing to agreed "problem" domains. In doing so, there is greater capacity for agency engagement through shared (individual and collective) understanding of outcomes and accountabilities.

Private sector

datanomics believe that the benefits of the "usage-driven" data sharing approach suggested for the public sector is also applicable to the private sector.

While large enterprises are already seeking to share data via various forms of protected “safe haven” models (eg Quantium - Woolworths, Qantas, NAB etc & Data Republic - Westpac, NAB, CBA etc), datanomics believe that the fostering of data sharing capabilities across the SMB sector offers the most significant private sector opportunity for economic and social value creation by encouraging industry collaboration and innovation.

The Australian economy is dominated by a relative few large enterprises across most major industry sectors. As such, these enterprises carry “natural” scale and resource advantages with respect to access

and use of their information assets. Policies that seek to influence the nature and level data availability need to mindful of whose interests will be served and whose may not.

An agile and resilient SMB sector is vital to Australia's economic and social prosperity. The empowerment of SMB's by providing meaningful access to the information economy through equitable data sharing models is important to "levelling the playing field".

Example: SMB data sharing

The following two SMB data sharing examples would not normally arise in a discussion about "data availability and use", but in the context of data sharing demonstrate how value creation already occurs:

ANZ Business Insights - the ANZ Bank provides all it merchants with access to comparative statistics that allow small business owners to track their performance against their peers over time and location. It combines payment data with industry classifications and customer demographics. The ANZ also use this data internally to inform economic analysis.

Stoploss Logic - are a Melbourne-based business that provides a subscription-based service that uses sophisticated algorithms to detect opportunities for retail loss prevention by analysing point-of-sale, inventory, and staffing data to identify anomalies, cost impacts and actions required. Comparative analysis across thousands of retail operations improves algorithmic effectiveness in detecting outlier events

Opportunity: Dynamic economic and industry data

The sharing of retail payments data could provide rolling retail and industry indexes at a level of granularity that could assist government policy and programs, researchers and the business sector.

The sharing of parcel delivery data could provide rolling insights about the impact of online purchasing by country (or state) of origin, product category, and delivery postcodes. This would be valuable for industry policy, and for SMB's in understanding local demand and online competition.

Both these examples also represent significant opportunities for data-driven innovation and value creation by providing both open and commercial information platforms.

Individuals and personal data

datanomics believe that the development of citizen/consumer side market infrastructures, that enable consumers to make better and informed consumption decisions offers the greatest potential economic and social value creation, however, is also the least mature or developed.

This view calls for a fundamental shift in thinking with regards to personal data. The World Economic Forum's multi-year Rethinking Personal Data initiative highlights the importance of moving from an institutionally oriented collection approach in which the individual is passive, to a shared data, shared governance model where trust and active engagement are possible.

Taken to an individual context, it is important to note that this embraces the same data sharing principles as discussed earlier in the context of public and private sector organisations.

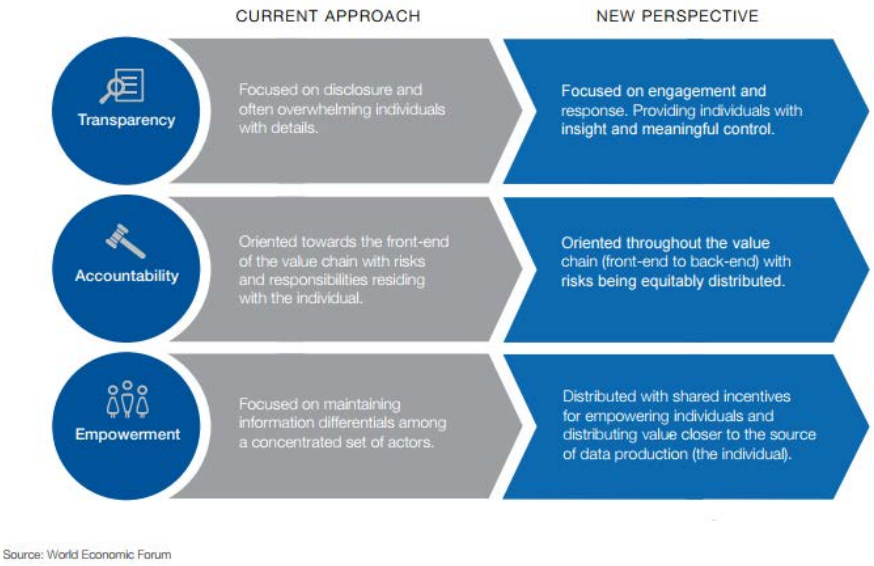
The following two diagrams from WEF reports by Boston Consulting Group (2013) and AT Kearney (2014) reinforce these points.

Comment: New Perspectives on the Use of Personal Data[5]:

---

[5] WEF Report Unlocking the Value of Personal Data: from Collection to Usage (Boston Consulting Group, 2013)

Comment: A New Lens for Strengthening Trust[6]:



Source: World Economic Forum

While no studies have yet been conducted in relation to the Australian economy, a CTRL-Shift (Jun-14) UK study[7] of the economic value of the Personal Information Management sector in the UK assessed its value at 1.2% of GVA or about 16.5b pounds. In crude comparative GVA terms, this might represent $15-20b in an Australian context.

Ctrl-Shift more recent report, "A New Paradigm for Personal Data" (June 2016) is instructive[8]. Commissioned by Facebook, the report is the output of a "roundtable series about how to create a positive and sustainable future for personal data - for the benefit of individuals, organisations, and societies", involving some 175 experts across 21 roundtables in 11 cities.

The report validates datanomics views with regards to the transformative effect that personal data will have across all sectors and industries, the immature and evolving nature of personal data ecosystems and fundamental importance if policies that support that balance personal, social and economic outcomes.

datanomics believe that strong policy leadership is key to providing citizens and consumers with rights to access, control and benefit from their personal data. This will, in turn, encourage the investment and innovation required to develop and build the requisite demand-side digital market capabilities.

[6] WEF Report - Rethinking Personal Data: A New Lens for Strengthening Trust (AT Kearney, 2014)
[7] Personal Information Management Services – an analysis of an emerging market (Ctrl-Shift, 2014) https://www.ctrl-shift.co.uk/insights/2014/06/16/personal-information-management-services-an-analysis-of-an-emerging-market
[8] A New Paradigm for Personal Data (Ctrl-Shift, 2016) https://www.facebook.com/anewdataparadigm

In data sharing terms, datanomics believe that the capacity for public and private sector organisations to engage with individual citizens or consumers through shared personal data has transformative implications for building trust and creating new forms of value exchange across all sectors.

In this personal data context, Harper review's recommendation about using data to better to inform consumer choice is a limited one. Rather datanomics' view, in line the WEF and Crtl-Shift reports, is that transformational social and economic value is not informed choice, but enabling informed consumption.

Informed choice is about the access to comparative product and service information. Informed consumption is about access to comparative individual consumption behaviour.

Opportunity: Australia Post Digital Mailbox

Australia Post has for some years been investing the development of a digital client platform in the form of its Digital Mailbox. While its initial focus has been on digital document and bill payment services, the platform could evolve into a personal data platform that could allow citizens and consumers to assume greater control over their personal data, and in doing so, open up new forms of value exchange with public sector agencies, businesses, and other consumers.

Comment: role of trusted intermediaries in personal data ecosystems

datanomics believe that the emergence of personal data as a transformative social and economic force is dependent on the development of demand-side actors, that by nature are trusted entities able to act with and on behalf of citizens and consumers. However, the growth of this sector will require greater regulatory support to ensure citizens have rights of access and control not present in current legislation.

For example, the EU GDPR, which will go into effect next makes strong provisions with regards to right of access to personal data, portability of format and penalties for non-compliance. In response, there has been significantly increased activity by both existing enterprises and new personal data start-ups to provide the type of trusted consumer-side applications and services that will enable them to benefit from their personal data.

Trusted data intermediaries widely exist, but most operate on closed data supply-side models that do not typically provide open access to data outside their respective ecosystems.

Dominant “sharing economy” business, such as Uber and Airbnb can be considered to operate as trusted intermediaries to match buyers and sellers. However, the “black-boxing” of processing the shared data at the heart of their business models is central to their ability to monetise their information assets and drive business valuation based on the expected future information value.

Data brokers and data exchange businesses such as Experian, Quantium, and more recently, Data Republic, offer trusted intermediary services by providing various forms of data "safe haven" capabilities. However, these are vendor and enterprise focused, not consumer side players.

Example: Meeco

A world leading example new consumer-side trusted intermediaries emerging the personal data space is Australian startup, Meeco (www.meeco.me). Founded in 2012, Meeco is recognised as one of the pioneers in its field. Their

application allows individuals to collect and manage personal data, but also leverage this information by being at able to engage with organisations in new ways. Meeco is currently focusing much of its activities in the UK and EU.

Standardisation

Standards are a means to an end, but not an end in themselves. The technology sector is replete with examples of attempts to create standards that have been made obsolete by market forces.

In terms of data, widely accepted standards exist offer significant value, however, caution needed on pursuing standards as a means to create shared value. This is typically the refrain of regulatory bodies. For opposite reasons, it is the refrain of incumbents seeking to protect their proprietary interests.

Data standards only represent part of standardisation challenge. Even where accepted data standards exist, a lack of technology and process standards may hinder access.

From a Usage-Driven perspective, the value creation opportunity may overcome a lack of standards.

Example: Thundermaps

Thundermaps, an NZ based startup has been accumulating potentially the world's largest open geospatial hazard database. In doing so, they have developed unique competencies in extracting open hazard data from agencies and organisations that do not have the capacity to do so for themselves.

Thundermaps' commercial model is to provide an occupational health and safety application to organisations that have field workers or contractors that may be exposed to hazards by providing hazard warning, logging and incident reporting. The EU is also using Thundermaps as a smart city tool for towns to provide similar hazard related capabilities that also include emergency services data.

In using Thundermaps, users also engage in updating and sharing hazard data thereby building shared value over time.

Even when the data is open and standards exist, many public and private organisations do not have the skills or resources to make data available. The Thundermaps example illustrates that understanding the value of "problem space" can drive innovation in data access rather than depending on its availability.

Ethics, Trust, and Privacy

Given the challenges that big data, IoT, and advanced analytics are placing on privacy and data protection matters, data ethics is a key issue that needs to be embraced. The Productivity Commission issues document addresses data ethics once, in relation to research ethics.

datanomics notes that there is significant global data ethics discourse on issues that include the following:

- Informed Consent - the ethics of how consent is obtained; ethical tests for what constitutes informed the consent; ethics of how unforeseen connections in data may impact consent; the ability for consent address undetermined future uses unknowable at the time of collection.
- Legitimate purpose and usage - the ethical considerations for determining legitimacy of purpose and usage; assessing for future purposes or usage that may be initially unknown and subject to discovery.
- Limitations on automated individual decisions and profiling – the ability to respect individual wishes not to be profiled or be identified as part of a group (ref EU GDPR).
- Algorithmic Accountability – ability to assess and monitor the ethical impact of algorithms that may result in unintended consequences over time.

Moreover, datanomics notes that over the past year there has been a significant global movement towards recognition of the importance of data ethics in policy and regulatory considerations of privacy, data protection, trust, and confidence. For example:

- The UK government has just announced (26/4/16) its intention to set up a Data Ethics Council[9].
- The US White House has just released a report (4/5/16) titled "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights" recognising the challenges and public risks posed by algorithmic systems[10].
- The European Data Protection Supervisor established an external Data Ethics Board (September 2015) to "better assess the ethical implications of how personal information is defined and used in the big data and artificial intelligence driven world."[11]

Data ethics provides for a much more encompassing framework for considering data across a wider range of usage settings. Considerations of what is private in a digital universe is increasingly a matter of context and personal control, and the scope of potentially personally identifiable data is expanding. To quote from the EU EDPS Ethics Board announcement about their purpose:

"...to reconsider the ethical dimension of the relationships between human rights, technology, markets and business models and their implications for the rights to privacy and data protection in the digital environment. …(and) identify a new ethical approach in the coming years so that individuals are no longer reduced to mere data subjects in the digital environment."

The work of the Information Accountability Foundation[12], an international policy think tank on "data protection law and practice through accountability-based information governance", have put suggested the following ethical values for data and advanced analytics:

Beneficial- both discovery and application phases (ie two-phase) require an organisation to define both the benefits and the parties that will benefit

Progressive - the value from data analytics should be materially better that might be achieved in a less data intensive manner

Sustainable - the effectiveness of algorithms and data over time changes and need to be taken into account

Respectful - relates to how the data originated, how it is to be used and the impact on all parties

[9] UK Government - Data Ethics Council announcement
[10] White House report - Big Data:A report on algorithmic systems, opportunity, and civil rights
[11] EU EDPS Data Ethics Board
[12] http://informationaccountability.org/

Fair - relates to the insights and applications that are the product of data

datanomics would also like to draw attention to a recent paper[13], which proposes a legitimate interests test to address regulatory shortcomings in relation to purpose limitation, consent and performance of an agreement concepts. While the paper relates to the EU GDPR regulations, datanomics believe it offers useful commentary relevant to this inquiry.

In summary, the paper proposes the following:

- That due to social trends and technological developments the principle of purpose limitation should be abandoned as a separate criterion.
- That other principles - such as consent and the performance of an agreement - should not be recognised as independent legal grounds to legitimise data processing
- That instead, a test based on whether there is a *legitimate interest for data collection and processing* (as well as further processing) of data should be applied.
- That such a test will provide for a more effective data protection regime that will have more legitimacy than the assessment under the existing legal regime that is primarily based on the purposes for which data may be collected and further used.

To highlight the current limitations of the purpose test and significance of an interests based one, the paper offers the following example:

Example: Purpose vs Interest

Let us imagine that during an evening out, an acquaintance mentions a mobile phone application that uses data relating to your movements and phone calls in order to inform you with a high degree of certainty whether you are likely to catch influenza. The app can even tell you which friends you should avoid in order to minimise your risk of catching the flu - even if those friends have not yet been affected by it themselves. Would you install this application on your smartphone as soon as you had the chance? Or would you prefer not to have your apparent future determined by data analysis?

This type of commercial service, with the capacity to build up a virtual picture of our state of health on a real-time basis, and which could even predict how we would be feeling the next day, would probably be a cause for consternation for some people. At the same time, however, the use of a very similar type of application by the World Health Organization (WHO), for example, could play a major role in protecting public health in specific cases of dangerous infectious diseases and pandemics. Two applications that both collect and process personal data for the same purpose: the monitoring and personalised prediction of health and illness. But the sentiments that these two applications would give rise to would likely be very different. The commercial application may not be welcomed with any great enthusiasm - at least not by everyone; however, the likelihood of societal acceptance would be much greater for an application that was used to contain pandemics to protect global health. And this would be true even if the WHO were to contract a commercial company to provide this service.

When we pause to reflect on this, we are bound to draw the conclusion that it is not so much the purposes for which personal data might be used that is the primary consideration here, but rather the interests that are served by the use of the data collected. And yet, the existing European Privacy Directive is based primarily on the purpose for which data are collected and processed, while the interests served play a much more subordinate role. This situation raises the question of whether the existing legal regime can be effective and can be considered legitimate as we move into a future that is driven by data.

---

[13] Moerel, Lokke, and Prins, Corien, Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things (May 25, 2016). SSRN: http://ssrn.com/abstract=2784123

The paper offers “5 pillars” in support of the need for a legitimate interest test. datanomics note that while terminology reflects the EU regulatory environment, datanomics believe there is applicability to Australia.

The first pillar - that in the past, personal data were invariably a by-product of the purpose for which these data were collected; however, technological developments entail that this is by no means always the case today.

The second pillar - that we must abandon the notion that parties that process data (in the terminology of the law: the data controllers) are acting lawfully simply by virtue of the fact that they notify individuals and ask them to click ‘OK’, when at the same time no careful consideration is made of the various interests that may be at stake with regard to the processing of those data.

The third pillar - that individuals become more and more transparent, except for themselves. The reality of today's data-driven society is that individuals often do not know which data are being processed about them, how they are being assessed and categorised by the data controllers, and what the consequences of this might be for them

[A relevant mitigating factor when evaluating the legitimate interest ground would be the extent to which the data controller provides individuals with effective control over their data.]

The fourth pillar - that the regime for special categories of personal data (health data, criminal data, religion, race and ethnic background, etc.), is no longer meaningful. Increasingly, it is upfront unclear whether data are sensitive. Rather, the focus should be on whether the use of such data is sensitive.

The fifth pillar - that the current system for assessing a data processing is too complex for the data controllers to apply. For example, under the existing system, a controller must demonstrate that a. there is an explicitly defined and legitimate purpose for its data collection and processing activities; b. that there is a legal ground for processing these data; c. that there is also a specific ground for processing special categories of personal data (and, if this test is less stringent than the test for regular data under b., then the processing must also meet the requirement under b.) and d. that any further processing of personal data is ‘not incompatible' with the original purpose for which the data were collected.

Suggested References in relation to Ethics, Privacy and Data Protection

datanomics recommend that following references offer useful sources of information provide on this matter:
Council for Big Data, Ethics, and Society: http://bdes.datasociety.net/

"In collaboration with the US National Science Foundation, the Council for Big Data, Ethics, and Society provides critical social and cultural perspectives on big data initiatives.

The Council brings together researchers from diverse disciplines — from anthropology and philosophy to economics and law — to address issues such as security, privacy, equality, and access in order to help guard against the repetition of known mistakes and inadequate preparation.

Through public commentary, events, white papers, and direct engagement with data analytics projects, the Council will develop frameworks to help researchers, practitioners, and the public understand the social, ethical, legal, and policy issues that underpin the big data phenomenon."

Professor Luciano Floridi
Director of Research and Professor of Philosophy and Ethics of Information
Oxford Internet Institute (OII) - http://www.oii.ox.ac.uk/people/floridi/
Luciano Floridi, Philosopher, Ethics of Information (OII): http://www.philosophyofinformation.net/

Floridi is best known for his foundational research on the Philosophy of Information and Information Ethics, two new research areas that he has contributed to establishing. He is also known for his work on the EU's Right to be forgotten (RTBF) legislation and is a member of the EU's Ethics Advisory Group on the ethical dimensions of data protection (which reports to the European Data Protection Supervisor).

Perspective - Privacy as public health issue:
Cory Doctorow's article http://boingboing.net/2016/07/03/peak-indifference-privacy-as.html

Perspective - Privacy as an environmental management issue:
"privacy injuries are to big data, as pollution is to industrial production…"[14]
Dennis D. Hirsch & Jonathan H. King, Big Data Sustainability: An Environmental Management Systems Analogy, 72 Wash. & Lee L. Rev. Online 406 (2016),
http://scholarlycommons.law.wlu.edu/wlulr-online/vol72/iss3/4

Crtl-Shift report on personal data for Facebook
Facebook - New data paradigm report https://www.facebook.com/anewdataparadigm

[14] Dennis D. Hirsch & Jonathan H. King, Big Data Sustainability: An Environmental Management Systems Analogy, 72 Wash. & Lee L. Rev. Online 406 (2016), http://scholarlycommons.law.wlu.edu/wlulr-online/vol72/iss3/4

datanomics

About datanomics

datanomics is a Melbourne-based data innovation company that develops and delivers self-governed shared data platforms for public, private and research sectors.

This submission was written by its founder and CEO, James Horton.