datanomics

response to draft report

James Horton - datanomics pty ltd

Productivity Commission Inquiry into Data Availability & Use

16th December 2016

Thank you for the opportunity to appear before the commission on 21st November.

This submission seeks to expand on issues discussed at the public hearing.

---

introduction

The commission is to be commended for its efforts in producing this draft report. To venture into such difficult and potentially unfamiliar domain, to gather and to make sense of such varied perspectives, and with equally varied and often undeclared agendas is an exceedingly difficult task.

In particular, the commission is to be commended on its recognition of the following:

- that in global terms, Australia has fallen to the "back of the pack"
- that fundamental reform is needed if we are as nation and society to equitably realise the value of our information assets
- the need to provide for individual rights over their data in the form of control and portability
- the need for both national regulatory reforms and the establishment of entities to facilitate appropriate accountabilities and governance in relation to data and information

However, my response to the draft report might be characterised as one of three "disappointments":

Firstly, that while a much needed new framework has been proposed, the report lacks an overarching narrative to provide a strategic context for its recommendations.

getting to the narrative: the first part of this submission addresses what might be incorporated in such a narrative.

Secondly, that the recommendations and proposed new framework are constrained by a data release frame of reference, that in the absence of a means of creating shared beneficial value, inherently fosters institutional risk minimisation and compliance behaviour.

getting to value - usage: the second part of the submission addresses the limitations of the proposed data release based framework and explores the value of a usage-based approach.

Thirdly, that the recommendations and proposed framework do not address the importance of data governance and data ethics in ensuring, not just the right accountabilities and practices, but morally sound outcomes.

getting to value - governance: the third part of the submission addresses role and importance of data governance and data ethics a national data framework.

Other comments and observations:

- institutional bias
- data sharing
- data ethics
- algorithmic accountability
- privacy and data portability
- open data

---

getting to the narrative

The issue of data availability and use is of enormous importance to everyone, everywhere, and most certainly much more so now than the authors of the terms of reference could have imaged even a year or so ago.

In that time we have seen, in the case of the ABS Census failure, the repercussions of poor data governance and management. While there was indeed a failure of technology, the far greater failure was one of trust, which was played out through the mismanagement of public engagement well before the census night event - the reputational cost of which will continue long after the technology costs are settled.

This should serve as a timely and valuable lesson on the importance of institutional and executive accountability, transparency, and data governance. However, the subsequent focus on technology and procurement failings suggest that the tendency to abrogate data governance responsibility to IT in the form of data management continues to prevail.

We have also seen how, despite the best efforts at anonymisation, released research data sets from Medicare were subject re-identification by University of Melbourne researchers. Concerningly, we also have the Federal Attorney General's attempt to criminalise reidentification of public data. That such legislation might even be considered is at best

symptomatic of a lack of subject matter expertise that potentially exposes us to substantial social and economic risks.

However, as will be explored later in this submission, the temptation to rely on technologists for " technology fixes" needs to be tempered by a much deeper conversation about expectations, values and responsibilities as they relate to our growing digital "personhood".

As the commission has rightly concluded, we are long past the usefulness of incrementalist approaches. However, meaningful and sustainable reform requires a new narrative, grounded not of ourselves in the context of an industrial society[1], but as part of an information one. An information society[2], where the role of technology is not that of enabling production as a basis for economic and social activity, but as a means of enabling the extraction, accumulation and manipulation of information for as a basis for economic and social activity.

This new information society environment, which brings together data, information, knowledge and communication is what Luciano Floridi[3] describes as the "infosphere"[4]. Like the concept of a biosphere, that brings together living things as part of an ecological system as a means of framing our thinking about our relationships and responsibilities within it, the concept infosphere offers an approach to reframe our thinking about relationships and responsibilities within an information society.

In the absence of higher level narrative, the draft report's attempt to offer a strategic perspective on the transformation of our data and digital landscapes succumbs to all too familiar and cliqued technology narrative about exponential data growth and enabling technologies. These descriptions offer little insight into the underlying drivers.

While an effective exploration of this subject area is beyond the time that it properly affords for this submission, the following brief perspectives are presented by way of examples:

- **the digital transformation of institutions and industry** is part of an accelerating arc that began over 50 years ago and has been the subject of much attention

    - for most of that time, the data landscape comprised data arising from operational systems and process from within organisations, including individual data in the form customer interactions
    - as consumers, we have been mostly aware of the data we provide to organisations in this context

---

[1] https://en.wikipedia.org/wiki/Industrial_society
[2] https://en.wikipedia.org/wiki/Information_society
[3] Professor of Philosophy and Ethics of Information, Oxford Internet Institute, University of Oxford
[4] Floridi - Ethics in the Infosphere - UNESCO speech (2001)

- **the digital transformation of individuals** has only occurred over a much shorter 10-15 year timeframe but has occurred with comparatively little attention

  - it is of a scope and scale that far exceeds the digital transformation of organisations
  - it has profound implications for our ideas about what it is to be human - such as identity, dignity, fairness and privacy - as much of thinking has been based on an "analogue" or physical sense of ourselves
  - that so much of about who we are and what we do exists in a digital form somewhere, and about which we have limited awareness or control, but if misused has the capacity to cause harm, raises fundamental questions

- **the emergence of the Big Data phenomenon** is significant, not for the scale of data, but for the fact that the underlying driver is data about our online behaviours (eg search, social, shopping)

  - big data is largely the result of the digital transformation of individuals - it is our "data exhaust"
  - for information economy organisations the value of this data lies in their ability to use it to influence our behaviours, which can be monetised through digital media and channel revenues
  - as individuals we much less aware of the data being collected or how is being used

- **the emergence of the Internet of Things (IoT) phenomenon** is significant, not for the number and scale of the things, but for the fact that the underlying driver is the "datafication" of the physical environments within which we interact

  - IoT creates a new form of "data exhaust", that of our environment and the things we use
  - one of the transformative effects of this datafication is to change the way physical assets can be consumed - from a capital asset to an "elastic" resource that can be consumed as service (eg software-as-a-service, cloud services, jet engine-as-service, car-as-a-service etc)
  - for information economy organisations the value of this data lies in their ability to use it to influence how we interact with our environment, which can be monetised through services
  - as individuals, we are even less aware and often unaware of the data being collected or how is being used

- **the rise of algorithms** has been fuelled by the need derive economic value from the data accumulations of information enterprises

  - algorithms can be thought of as the industrial robots of the information economy, and whose capacities for autonomous behaviours are growing as learning capabilities grow

- - while most algorithmic techniques (eg neural networks and machine learning) are not new, their rapid development and proliferation has been enabled by the availability of sufficiently large data sets to train them and the reduction in the computing costs to run them
    - one of the most significant issues associated with these advanced algorithmic technologies is that their underlying complexity renders them opaque even to those using them, and in the case of neural networks, inscrutable[5] - which, notwithstanding the potential value they might promise, gives rise to major ethical and accountability concerns

- **the emergence of virtual and augmented reality technologies**, once considered the domain of gaming and entertainment industries, are being driven by perceived economic opportunities to be realised by harnessing digitally transformed individuals and environments as part of virtual and blended (augmented) reality ones

- in this post-industrial era, the big question that overshadows all others in relation to the availability and use of data is this - **how do humans live well in an information society?**
    - the most profound difference between industrial and information economic and social landscapes is this:
        - in an industrial society, power arises from the capacity to control the factors of production
        - in an information, society power arises from the capacity to influence the factors of behaviour modification[6]
    - the responsibilities, rules and institutions that have arisen to ensure an industrial society that is fair and equitable one for humans needs to be fundamentally reconsidered for an information society

These perspectives offer an indication of the level of context that might support to a narrative capable of framing the key elements of the inquiry's terms of reference.

---

recommendations:

revise the report overview to communicate a narrative capable of:

- placing the key requirements of the terms of reference within a more cohesive context;
- providing a more coherent basis for integrating and assessing the report, and;
- providing a strategic context for the proposed framework

---

[5] Bornstein - Issue 40 - Nautilus - Is artificial intelligence permanently inscrutable? (Sep-2016)
[6] Zuboff - Big Other: Surveillance Capitalism and the Prospects of an Information Civilization (2016)

getting to value - usage

Without a conceptual basis for understanding how the informational value of data is realised, the draft report is unfortunately limited to relating case studies and reports. Furthermore, the report's underlying "data release" frame of reference results in a proposed framework that relies on establishing new bureaucratic and regulatory mechanisms to facilitate release.

The "data release" frame of reference that is embedded in both the inquiry's terms of reference and the subsequent draft report presents a major impediment to any meaningful reform for the following reasons:

- seems to offer little to no mechanism for encouraging collaborative data sharing behaviour
- appears to do little to prevent the continuation of risk avoidance and compliance behaviour;
- creates a new level of bureaucracy that may potentially serve to distribute risk and encourage regulatory outsourcing of risk rather than maintaining accountability

While to some extent this has been imposed by the implicit assumptions underpinning the inquiry's terms of reference, the commission could accommodate this by applying a higher level of abstraction or conceptual frame of reference, rather than taking the terms of reference as a literal starting point.

A more useful conceptual starting point is the recognition that:

1. data has no intrinsic value; and,
2. the value of data is based on the *informational value realised on usage*

And that this usage value manifests itself in two ways:

1. **the potential for harmful outcomes to arise from the use of the data**
   - these harms might be in relation to a personal, group, community or national nature and could relate to matters such as privacy, discrimination, health, law, law enforcement or national security
   - there may also be harms of an economic or environmental nature that might impact, not only to individuals but organisations
   - typically such data is characterised by the nature of it origins and usually subject to some form of regulatory considerations

2. **the potential for beneficial outcomes to arise from the use of the data**
   - these benefits may be of economic, social or environmental nature and relate to individuals, groups or organisations
   - typically such data is characterised by the nature of its usage and subject to the legitimate interests and purpose associated with the usage domain

On the basis of this approach, two types of "accredited authorities" might more usefully contribute to the proposed framework:

1. **Data origination focus - "Data Access Authorities"**

    ○ would address "special interest" data categories and sources, that need to be governed and managed in a particular way, potentially throughout the data life cycle, in order to minimise the potential for harms to arise

    ○ focused on the specialised needs of specific data origination domains, these authorities would require specific capabilities and competencies to manage the associated risk and compliance issues

    ○ would be closely associated with their respective data origination domain - typically institutionally based, though not exclusively so

    ○ would be of particular value in focusing capabilities where multiple sources of sensitive data may be required to be brought together

    ○ would also provide a valuable aggregation point of specialist expertise and data access for research needs where there exist additional research ethics considerations

    ○ might also offer the potential for developing into centres of expertise, that could contribute, as part of a national network, towards the development of new data policies, guidelines and standards

---

"Special interest" data categories can be readily identified (eg regulation) and offer a much more meaningful (and measurable) basis for authorisation than the data types contained in the proposed framework. (ref figure 3 draft report)

---

example:

Data sharing across police, legal, health and community services agencies as part of the Victorian Government Family Violence Plan

---

2. Data usage focused - Data Usage Authorities

- would address "shared data usage domains" or problem spaces that are broadly defined by the economic, social or environmental value they address

- would be formed on the basis of addressing problems or opportunities that can only be realised through the sharing of data

- established on an independent, shared governance, shared data basis

  - comprising representatives of the data contributing organisations
  - leveraging principles of reciprocity, shared value, and shared accountability to build trust and encourage collaboration
  - issues of data release with regards to commercialisation, open data, or research would be addressed by the shared governance entity

- in a public sector context, they would offer a shared data value-pull model as opposed to a release-push one, with usage domains being cross-agency led or based on existing policy priorities

- the life cycle of these "authorities" would be determined by the shared value of usage domain

---

"Shared usage domains" can more readily be defined by the nature of the shared problem space, the associated shared economic and social value, as well as associated data contributors. They also offer a much more meaningful (and measurable) basis for authorisation than the data uses contained in the proposed framework (ref figure 3 of draft report).

---

public-private-research example:

A data sharing initiative to address national infrastructure resilience[7].

Brings together major utilities, energy, communications, essential services, local, state and federal government, and research to identify infrastructure resilience investments, ensure resilience criteria are factored into infrastructure projects, and develop more effective shared response and recovery plans etc.

---

[7] http://australianbusinessroundtable.com.au/

Enables collaborating via shared data usage domain based on a shared value proposition comprising:

- the economic cost of natural disasters - currently $9b annually rising to $33b by 20050
- the 10:1 investment disparity between spending response-recovery and resilience mitigation - also reflected in research, and
- an estimated $12.2b of budget savings that could be realised through approximately $250m of targeted infrastructure investment

public service example:

In Victoria there are already a number of current data sharing policy or program commitments:

- as part of Family Violence (as mentioned above)
- the Centre for Data Insights (under DPC), and
- a shared information agency as part of the DHHS' Towards Zero Safety plan

In this data value context, the terms of reference driven need to focus on identifying a wide range high value and national interest data sets is potentially reduced to initially focusing on where they might align with special interest data origination domains that exhibit high harm potential.

While it does not obviate the high value and national interest requirements, a shared value led "accredited usage authority" approach offers the potential for a building a more dynamic shared data, shared governance ecosystem on a national scale.

recommendations:

incorporate a data value framework that:

- recognises values based on the potential for harm and benefit

- aligns accredited authorities with origination and usage domains

- allows for the development of harmonised data governance-led approaches to the provisioning of "special interest" data origination domains

- replace the concept of Accredited Release Authorities with the following:

- ○ Accredited Data Access Authorities

  - ■ accountable for the administration of "special interest" data origination domains
  - ■ maybe institution or program based
  - ■ maintain a whole of data life-cycle risk and harm minimisation focus
  - ■ network with other Data Access Authorities
  - ■ develop and deliver educative and advisory services
  - ■ act as a point of engagement for research and industry

- ○ Accredited Data Usage Authorities

  - ■ shared usage domain defined by shared problem space
    - ● data contributors engage on the basis of usage domain requirements
    - ● problem space defined by the participants
  - ■ governed independently on a shared governance basis by its data contributor organisations
  - ■ may exist within an institution as cross division/group shared governance entity
  - ■ typically established as a cross agency or cross-sector public-private-research shared governance entity

---

It may also be useful to note that at a data level of abstraction, there is a further value model.

Value (harm and benefit) can be enabled in a number of ways, by:

- ● data accumulation over time and space
- ● discovering new usage opportunities for existing data sources
- ● enabling data to be combined across multiple sources
- ● generating new derived data arising from analysis
- ● enabling new forms of value exchange between interested parties
- ● encouraging collaborative behaviours that realise process and resource sharing value

---

getting to value - governance

The draft report is constrained in its capacity to address issues of availability and usage by an implicit "data release" frame of reference. This manifests itself in two ways:

Firstly, a data release orientation results in a tendency to focus on assessing the cost-effectiveness technology and data management issues associated with minimising post-release risks (eg anonymisation, data linking, standards, API's etc).

Without a notion of usage value in the proposed framework, cost and risk become the focus.

While the proposed framework does need to offer a capacity to release data in a standardised manner to standardised usage environments (as illustrated in the draft report), the lack of a value concept in relation to usage means that the framework reduces to being compliance and "cost of output" driven, without a "benefit of outcome" driver to encourage collaborative behaviour.

The suggested redefining of Accredited Release Authorities with "special interest" data origin focused "Access Authorities" and the inclusion of a new shared data usage focused "Usage Authority" provides a means addressing these issues.

Secondly, focusing on the process of releasing data tends to a view of data governance as a data management function.

In other words, that data is viewed as being "physicalised" as content within a technology infrastructure, and data management processes (such as anonymisation, linking, and publishing) are an IT responsibility and by association, data governance.

---

examples:

- that data release risks can be adequately managed via technological means actions such as de-identification, obfuscation or encryption

- that increased data usage does not lead to increased risk because the right application of technology and regulation will be sufficient to protect individual and institutional interests from harm.

The key principle is this:

> *Data management is a function of data technology infrastructures that enables it to give effect to the data governance requirements of those who are accountable for it.*

Data governance comes before data management. While business, operational and technology units should carry data governance responsibilities, strategic accountability it should ultimately reside at the most senior executive or board levels.

However, the reality for almost all public and private sector institutions is that data management and data governance are conflated and accountability for data governance abrogated to operational and technology management. This has been so because data has traditionally been seen as being generated, managed and contained within an organisation's core operational systems along with the technology infrastructures upon which they run.

In other words, control the technology and you control the data. To the extent that data might have been regarded as an asset, it has historically been sufficient to regard its value as being embedded in the application that created it.

This is a reality that is fading rapidly. Not only is data is now generated by much more than an organisation's operational systems, but the forces of digital transformation mean that the physical "boundaries" which once defined their internal and external interactions have now given way to more porous digital ones. Data that once were managed within internal systems, now flow across applications that have long since been abstracted away from their technology infrastructures, which themselves have been virtualised and reborn as services.

The fundamental data governance issue then, is this:

> *If data is to be regarded as-as asset then it needs to be governed as one and afforded the same level of attention as applies to the governance of an organisation's physical, human, operational and financial assets.*

If the value of data and information assets were explicitly represented on the financial reports of our major institutions and afforded the same board level attention as physical assets, there would little argument about accountability. However, it is also true that if governing data were that simple, it would be common practice now.

Difficult or not, to embark on a national data reform such as this, without reference to the role of data governance, would risk a the implementing a system that, with the best of intentions, is likely to become inexorably stymied by the dynamics of a technocratic bureaucracy pitted against the risk avoidance and compliance behaviours of agency leadership.

On the upside, the proposed framework will certainly gain the support of existing open data initiatives seeking some form of revitalisation, and it will more than likely attract plenty of in-principle support for the other data accessibility scenarios, along with the proposed structural and regulatory reforms. However, commitment to reform may well founder upon the lukewarm embrace of the status quo[8].

The success of a national framework such as this, however, is within reach, but it hinges upon one the realisation of one particular reform:

> *The securing of public sector wide acceptance that executive leadership (and/or board) has strategic accountability for the governance of its data.*

Of course, it would be naive to think that agency leadership would be eagerly lining up to sign on, particularly given that the idea of strategic accountability for the governance of data at an institutional level (as distinct from division, group or information technology related levels) is relatively new - and even more so outside of compliance related accountabilities[9].

However, getting beyond ad-hoc and tactical outcomes to unlocking the strategic value of the public sector's data assets calls for systemic change that can only be achieved through collective commitment and actions of its leadership.

> *For leadership to engage, the main drivers would need to be about the opportunity for both individual and cross-agency value creation (and before risk and compliance).*

---

[8] hacking Machiavelli's quote on the perils of the innovator and the pursuit of a new order of things
[9] note: in some industries executive level data governance accountabilities currently exist in the form of regulatory compliance obligations, such Basel II, Sarbanes-Oxley etc

Newness may well hold a key to the success of this reform by providing a way for public sector leaders to co-develop their strategic data governance guidelines via an opt-in program. In doing so, the program would seek to be a catalyst for change, hopefully of the transformative kind. (ref: appendix: A)

---

recommendations:

- include in the final report, reference to and recognition of the role of strategic data governance in supporting:

    - the functions of accredited authorities
    - public sector executive leadership engagement

- incorporate the development and promotion of data governance with the proposed peak organisation (ie NDC in the proposed framework)

    - note: with reference to the establishment of a data ethics council, host this under the NDC

- consider changing the name of the NDC to Data Australia or the National Data Australia to reflect an expanded role with respect to carrying data governance and data ethics responsibilities

- refer appendix B: suggested framework

Other comments and observations

institutional bias

Notwithstanding the issues relating to individual control of data, the draft report is institutionally biased and does not address the impacts on smaller organisations, charities, not-for-profits and NGO's that dominate the domestic economic landscape. The organisations are at risk of being marginalised in a transforming information economy by larger organisations seeking to maintain their market power. While this may in part be due to the terms of reference, it is nonetheless an important issue.

Approaches to data sharing and governance offer potentially transformative opportunities for sectors such as charities, not-for-profit, and creative and primary industries. This type of approach could address shared market and industry problems by enabling more informed supply-side and demand-side capabilities.

recommendation:

- encourage the establishment of shared governance, data sharing "AUA's" as an economic and industry development strategy for SMB intensive sectors

data sharing

The term data sharing is used widely in the report, often in the context of a data release process. In many instances in the draft report what is being described as data sharing is actually data exchange.

data sharing is not data exchange

If data rights are relinquished on data release it is data exchange (and not data sharing).

If data is shared then the data contributor (or release authority) retains a right over the data (solely or jointly).

---

data ethics

Page 314 was disappointing, particularly the paragraph relating to data ethics…

> *"Finally, some participants have noted that data ethics codes can play a role in driving better practices across industry (Australia Post, sub. 174; Datanomics, sub. 129). While these are positive developments, it is not clear to the Commission that the government has a role to play here."*

It should be noted that Australia Post and datanomics did not actually discuss data ethics in the context of "ethics codes" or "driving better practices", but rather the need to recognise the importance of data ethics as a new branch of ethics (and quite separate and distinct from the research ethics) that has arisen to address the challenges brought about phenomenon such as big data, IoT and advanced analytics.

While data ethics codes of practice are among the possible outcomes, the primary purpose of bringing data ethics into the frame of this inquiry is to bring critical new thinking to a (data) domain that is fundamentally being transformed, and where the prevailing discourse has been dominated by too few the voices - that of technologists and regulators.

Given that all of the 80 references to ethics, but this one were in reference to research ethics, it understandable that the inquiry may have misconstrued our reference to data ethics. Data ethics has a vital role to play in the well-being and sustainability of a national data framework. Its inclusion will not only be an important a catalyst developing much needed new knowledge and competencies but will provide a basis for engagement with recently established national data ethics initiatives in the UK, EU and US.

The following description of data ethics, along with some selected references should sufficiently make the case for its value, importance and inclusion.

What is Data Ethics?[10]

Data ethics is emerging as new branch of ethics that seeks to study and evaluate the moral problems related to:

---

[10] https://www.academia.edu/30234860/What_is_Data_Ethics

The ethics of data. The definition of ethical principles ensuring fair data handling and protection of individual rights while using large datasets for scientific or commercial purposes.

The ethics of algorithms. The analysis of ethical problems and of the responsibilities and accountabilities of designers and data scientists with respect to unforeseen and undesired consequences as well as missed opportunities concerning the design and deployment of complex autonomous algorithms.

The ethics of practices. The identification of the appropriate ethical framework to shape a deontological code about responsible innovation and data management to ensure ethical practices fostering both the progress of data science and the protection of the rights of data subjects.

What sort of issues does data ethics seek to address[11]?

The 5 main areas of concern data ethics seek to address are as follows:

1. privacy (including anonymisation and data protection)
2. informed consent
3. data ownership (the right to control and the right to benefit)
4. epistemology and objectivity (ie the opacity of advanced algorithms)
5. digital divide (ie social discrimination, inequalities and inequities)

In addition, there are a number emerging areas of concern and research. These include:

- the dangers of group-level ethical harms (group privacy and the right not to categorised as being part of a group)
- the changing nature of fiduciary relationships that become increasingly data-saturated
- the need to distinguish between "academic" and "commercial" (big) data practices in terms of potential harms to individuals
- future problems with ownership of intellectual property generated from the analysis of aggregated datasets
- the difficulty of providing meaningful access rights to individual data subjects

---

[11]
https://www.academia.edu/12611054/The_Ethics_of_Big_Data_Current_and_Foreseeable_Issues_in_Biomedical_Contexts

Selected data ethics quotes in relation to UK, EU and US national initiatives:

UK (2016): The UK government announced the establishment of a Council of Data Ethics...

> *"as a means of addressing the growing legal and ethical challenges associated with balancing privacy, anonymisation, security and public benefit."*

EU (2015) - the European Data Protection Supervisor announced the establishment of a Data Ethics Board in order to…

> *"help define a new digital ethics, allowing the EU to realise the benefits of technology for society and the economy in a way that reinforces the rights and freedoms of individuals".*

US (2014) - Supported by the US Government via the National Science Foundation, the Council for Big Data, Ethics, and Society was established...

> *"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."*

> *"...the Council develops "frameworks to help researchers, practitioners, and the public understand the social, ethical, legal, and policy issues that underpin the big data phenomenon."*

---

recommendations:

- establish a national data ethics group as part of the proposed national framework

- convene the group under the auspices of the NDC as the peak data organisation

- use the group:
    - to bring together leading thinkers from diverse disciplines such anthropology and philosophy to economics and law to develop address issues such as privacy, equality, access, anonymisation, security and public benefit (ref UK, EU and US initiatives)
    - as means of engaging civil society organisations, such as the Australian Privacy Foundation, Internet Australia, Open Knowledge Foundation etc
    - as a basis for engagement with national data ethics councils in UK, EU & US

- refer appendix B: suggested framework

---

data ethics additional references:

Of particular note are the following Accenture Labs publications, authored by Ethical Resolve, a data ethics research and consulting firm (published earlier this year):

Building Digital Trust: The Role of Data Ethics in the Digital Age
Universal Principles of Data Ethics: 12 Guidelines for Developing Ethics Codes
Facilitating ethical decisions throughout the data supply chain
Ethical algorithms for "sense and respond" systems
Informed consent and data in motion

The Ethics of Data Sharing: A guide to best practices and governance

The latter report on the "Ethics of Data Sharing: A guide to best practices and governance" is particularly pertinent to the commission's terms of reference.

---

extract:

Twelve Principles of Data Ethics (Ethical Resolve, June 2016)A set of universal principles of data ethics can help guide data science professionals and practitioners in creating a code of data ethics that is specific and contextual for their organisation or community of stakeholders: **1. The highest priority is to respect the persons behind the data.**

> Where insights derived from data could impact the human condition, the potential harm to individuals and communities should be the paramount consideration. Big data can produce compelling insights into populations, but those same insights can be used to unfairly limit an individual's possibilities.

2. Account for the downstream uses of datasets.

Data professionals should strive to use data in ways that are consistent with the intentions and understanding of the disclosing party. Many regulations govern datasets on the basis of the status of the data: "public," "private" or "proprietary", for example. But what is done with datasets is ultimately more consequential to subjects/users than the type of data or the context in which it is collected. Correlative use of repurposed data in research and industry represents the greatest promise and the greatest risk of data analytics.

3. The consequences of utilising data and analytical tools today are shaped by how they've been used in the past.

There's no such thing as raw data. All datasets and accompanying analytic tools carry a history of human decision-making. As far as possible, that history should be auditable. This should include mechanisms for tracking the context of collection, methods of consent, chains of responsibility, and assessments of data quality and accuracy.

4. Seek to match privacy and security safeguards with privacy and security expectations.

Data subjects hold a range of expectations about the privacy and security of their data. These expectations are often context-dependent. Designers and data professionals should give due consideration to those expectations and align safeguards and expectations with them, as much as possible.

5. Always follow the law, but understand that the law is often a minimum bar.

Digital transformations have become a standard evolutionary path for businesses and governments. However, because laws have largely failed to keep up with the pace of digital innovation and change, existing regulations are often miscalibrated to current risks. In this context, compliance means complacency. To excel in data ethics, leaders must define their own compliance frameworks to outperform legislated requirements.

6. Be wary of collecting data just for the sake of having more data.
The power and peril of data analytics are that data collected today will be useful for unpredictable purposes in the future. Give due consideration to the possibility that less data may result in both better analysis and less risk.
7. Data can be a tool of both inclusion and exclusion.

While everyone should have access to the social and economic benefits of data, not everyone is equally impacted by the processes of data collection, correlation, and

prediction. Data professionals should strive to mitigate the disparate impacts of their products and listen to the concerns of affected communities.

8. As far as possible, explain methods for analysis and marketing to data disclosers.
Maximising transparency at the point of data collection can minimise the more significant risks that arise as data travels through the data supply chain.

9. Data scientists and practitioners should accurately represent their qualifications (and limits to their expertise), adhere to professional standards, and strive for peer accountability.

The long-term success of this discipline depends on public and client trust. Data professionals should develop practices for holding themselves and their peers accountable to shared standards.

10. Design practices that incorporate transparency, configurability, accountability and auditability.

Not all ethical dilemmas have design solutions, but paying close attention to design practices can break down many of the practical barriers that stand in the way of shared, robust ethical standards. Data ethics is an engineering challenge worthy of the best minds in the field.

11. Products and research practices should be subject to internal (and potentially external) ethical review.
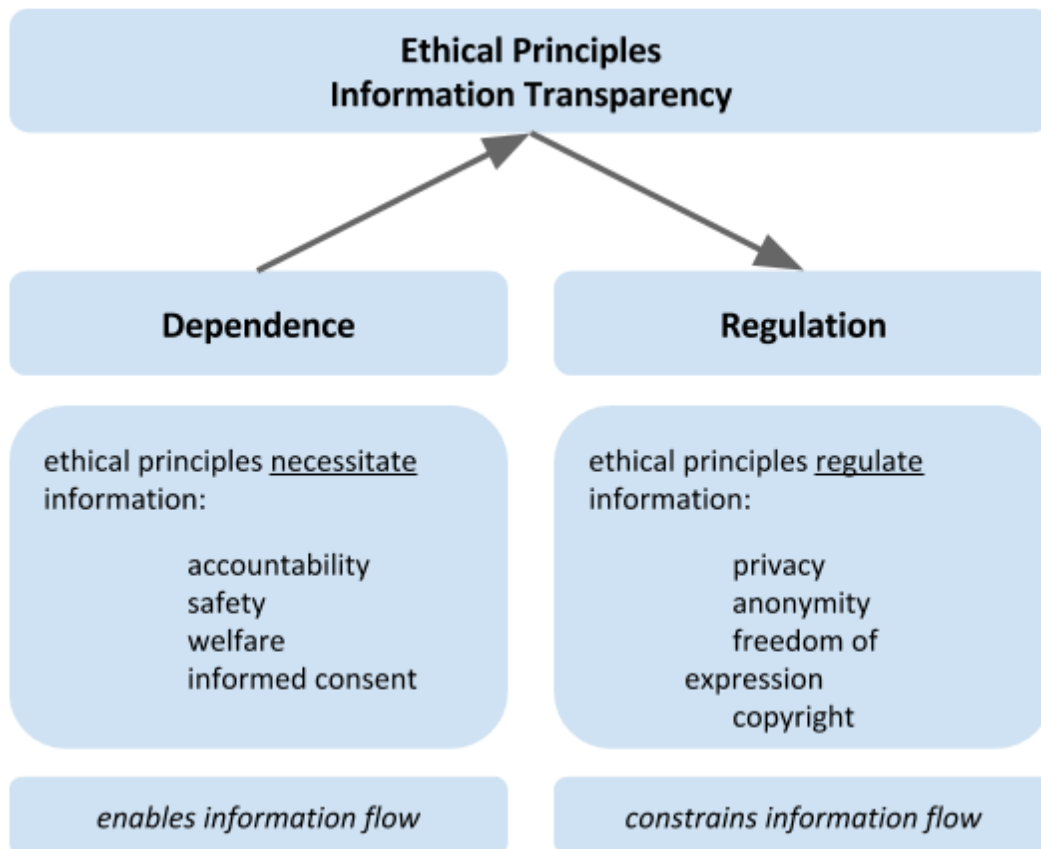
Organisations should prioritise establishing consistent, efficient and actionable ethics review practices for new products, services and research programs. Internal peer-review practices help to mitigate risk, and an external review board can contribute significantly to public trust.

12. Governance practices should be robust, known to all team members and regularly reviewed.

Data ethics poses organisational challenges that can't be resolved by compliance regimes alone. Because the regulatory, social and engineering terrains are in flux, organisations engaged in data analytics need collaborative, routine and transparent practices for ethical governance.

extract:

[Oxford Internet Institute - The ethics of information transparency](#) (Turilli, Floridi 2009)

the sociopath in the code - algorithmic accountability

One of the provocations I often use with data scientists is to consider what the psychological profile of an advanced algorithm it might be.

A: it is a sociopath

Empathy free, and perfectly responsive to the data it encounters in accordance with the logic of its design. Unconcerned about bias or shifts in data over time, and that were potentially

unforeseen by its maker, and equally unconcerned by the consequences of its inherited and learned design biases.

The issue of algorithmic accountability represents one of the most important justifications for data ethics. The use of algorithms is exploding. The industrial robots of our information society, they are created and operate largely unfettered by regulation or oversight, and with little accountability. Moreover, they operate in the main as "black boxes" often with little to no understanding of how and why (in the case of neural networks) they work.

We do not need to look far to understand their influence on world events, but of equal concern should their invisible influence on our access to the goods and services we use.

One of the consequences of the data release or distribution orientation of the framework proposed by the commission, are the underlying assumptions that accountability and governance are transferred on release; and, that provided data is made available in a suitably de-identified form, and/or handled in particularly secure manner, the data contributor is largely absolved from the how the data may be used.

This regulatory compartmentalising of accountability and risk both diminishes the capacity to derive social and economic value, as well as the capacity to ethically assess the impact of advanced algorithmic processes on data interest parties.
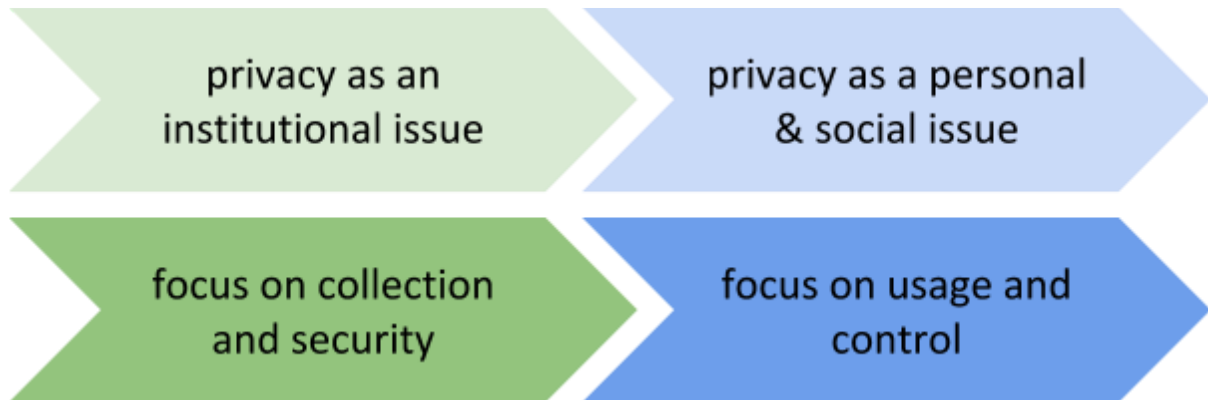
---

recommendation:

- research and development of ethical frameworks for the design and oversight of algorithms should be a priority for a data ethics council

---

privacy and data portability

While the draft report's recommendations to provide individuals with greater control of their data is to be commended, this demands a profound change in the way both public and private institutions think about privacy.

The proposed regulatory reforms understate the complexity and difficulty of this undertaking.

In conceptual terms, this requires a shift from thinking about privacy as an institutional issue to one that is a personal and social issue:

privacy as an institutional issue	privacy as a personal & social issue
focus on collection and security	focus on usage and control

Regulatory reform is critical to give substance to individual rights, but a much bigger shift is required on the part of individuals.

Unless regulatory reforms provide both accessible remedies as well as substantive remedies in law, then any rights to control data are pointless. This similarly applies to existing privacy legislation.

note: the introduction of occupational health and safety legislation offers a useful example of lack legislative effect until accessible remedies are available.

Beyond legislative reform to encourage institutional change, the notion that there is pent-up demand from individuals ready to "take control" of their data is a naive one. The reality is that decades of "supply-side" institutional information control and power over customer data will not change without new "demand-side" consumer advocate organisations capable of building trusted relationships with individuals, and can leverage their data to deliver new forms of mutual value.

For these new "demand-side" organisations, as well as their legacy "supply-side" counterparts, this means a fundamental shift from viewing privacy as a compliance "destination" to privacy as a foundation for building new value through trust.

For individuals, who are in the main disempowered, "privacy poor" and have limited awareness about their digital personhood, the challenge is even greater. The privacy issue here can not be resolved by regulation. This is more a matter of developing an understanding of "digital hygiene" and of education.

In this regard, *the issue of privacy might better viewed as a public health issue*[12].

Providing individuals with control of personal data is comparatively straightforward. The challenges are in resolving matters of identity and what constitutes personal data. Both of which are contextually fluid concepts. Even given the resolution of these issues, the biggest transformation is the attitudinal shift is required on the part of individuals about assuming the level accountability that accompanies control over their data.

Another important privacy-related issue absent in the draft report is the impact of Australia's lack of a Bill of Rights. As the only western democracy without one, its absence potentially leaves us increasingly more vulnerable to digital personhood harms by limiting our ability to seek redress and by limiting legislation that may seek to diminish those rights.

---

recommendations:

- Encourage public sector and civil society to reframe privacy as a "public health" issue.
- Encourage debate about the necessary constitutional or legislative provisions needed to provide individual rights protections that encompass digital "personhood".

---

open data

The issue of Open Data is worth commenting on given its absence in the narrative related points explored earlier.

The rise of open data is significant insofar as it is a movement, rather than a technological phenomenon, and whose origins have arisen out of the ideologies of open source, open government and open access movements of the early to mid-2000's. Accordingly, its technology proponents have tended to be drawn from the web and application community rather than data management and analysis. This is reflected in a tendency to treat data as content to published in the form of datasets and exposed via web portals. Similarly, data linking, as a data access method finds its origins in record and (web) page linking paradigms.

While the open data movement has grown out the universality of the world wide web and ready access to open source technologies, its relationship with the underlying data and information drivers for big data and IoT phenomenon are relatively superficial.

---

[12] [Doctorow - Locus Jul-2106 - Peak Indifference - privacy as public health](#)

The economic and social value of open data has largely been associated with open government, transparency, and civic engagement activities (eg hackathons, competitions etc). Applications arising from these efforts are mostly process oriented, mobile, often geospatial in some form, and in data terms, descriptive rather than analytical.

Optimistic analyst predictions about the economic value of open data, which seem to have peaked some 3-4 years ago, have not materialised and there appears to be an emerging discourse about the need to better understand the nature of usage and measurement.[13]

In the context of this inquiry, the "data release" and "data as content" ideologies that are central tenets of the open data movement continue to maintain significant inertia within governments worldwide, including their influence on this inquiry's terms of reference.

---

[13]

http://opendatacon.org/how-to-advance-open-data-research-towards-an-understanding-of-demand-users-and-key-data/

Appendix A: example leadership program

A suggested program might seek to achieve the following:

- enable participants to develop a level of strategic awareness and understanding about the nature of their respective agency's information and data assets
  - note: most likely this will be for the first time and hopefully a transformative knowledge
- using newly acquired data asset knowledge, identify potential data-sharing usage domains that represent strategic shared problem/opportunity spaces (intra- and/or inter-agency)
  - note: these shared data usage domains could form the basis for the suggested "usage authorities"
  - the object of this exercise is to set up a focus on the value creation and establish a compelling incentive for action
- on the basis of their newly acquired data asset knowledge, and in the context of identified shared usage domain (ie problem/opportunity space), collaboratively develop draft data ethics principles that reflect the value and principles required to support the usage domain
- on the basis of their newly acquired data asset knowledge, collaboratively develop draft strategic data governance guidelines that give effect to the draft data ethics principles for individual agencies, as well as shared governance guideline as they might apply to the identified shared usage domain.

Appendix B: suggested framework

The diagram shows a framework with:

- **data origin** → accredited access authorities
- **data usage** → accredited usage authorities
- **NDC**

- **data ethics** — eg values, legitimate interests, dignity, fairness, consent
- **data governance** *(gives effect to data ethics requirements)* — eg privacy, accountability, trust-authorisation, rights, risk, compliance
- **data management** *(gives effect to data governance requirements)* — eg identity, anonymisation, data linking, security, quality, provenance, publication