

Data Availability and Use

Submission to the Productivity Commission

Data Republic Pty Ltd

Level 1, 50 Bridge Street, Sydney

August 2016

Contents

1. The case for data sharing	3
1.1. What is data sharing?.....	3
1.2. Economic and social outcomes of data sharing	3
1.3. Assessing the opportunity and benefits.....	3
1.4. Market need for data sharing.....	4
2. Introduction to Data Republic.....	5
2.1. Our business objective	5
2.2. Our experience over the last two years	5
3. Risks and issues related to data sharing	6
3.1. Governance	6
3.2. Privacy guidelines adherence.....	6
3.3. Risk of identification of personal information	7
3.4. Legal negotiations	7
3.5. Time and cost to implement	8
3.6. Risk of re-identification of personal information.....	8
3.7. Security.....	8
3.8. Safe, neutral harbour for data.....	9
3.9. Pricing and valuing data	9
3.10. Lineage.....	9
3.11. National security and fraud.....	10
3.12. Giving consumers control of their data	10
3.13. Data exchange is only possible for large organisations	10
3.14. Data Sovereignty	11
4. Background on Data Republic.....	12
4.1. Company overview	12
4.2. Data Republic's ecosystem	12
4.3. The future	12
4.4. Company structure	12

1. The case for data sharing

1.1. What is data sharing?

Data sharing occurs when one organisation enables another organisation to access its data. The data is effectively ‘shared’ with the other organisation. ‘Organisation’ could refer to companies, governments or not-for-profits.

Many organisations possess data that would be of value to other organisations. However, these data resources are usually tightly held such that it is difficult, if not impossible, to share in a secure manner.

Data sharing is an essential activity in increasing the availability and use of data, as per the focus of this inquiry. The Australian economy does not suffer from a lack of data. It suffers from a lack of liquidity in data and therefore data is under-utilised as an asset class.

1.2. Economic and social outcomes of data sharing

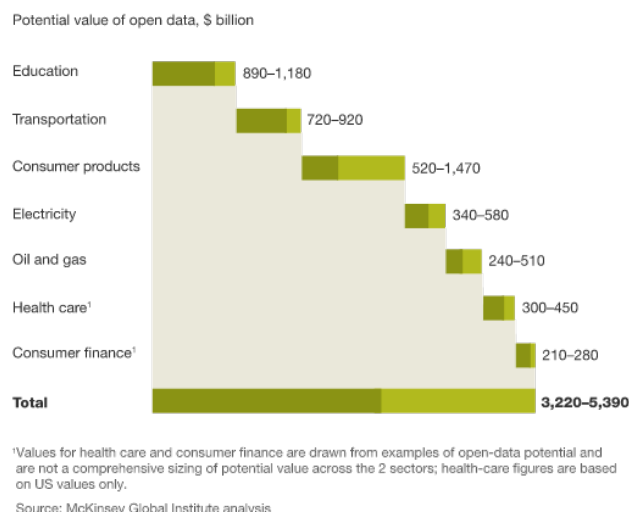
The Productivity Commission’s Terms of Reference for this inquiry outline the potential benefits of increased availability and use of data arising from data sharing. We strongly agree with the outlined points.

We believe that the primary outcome of greater availability of data is to enable better decision-making. This results in superior allocation of resources, increasing efficiency and effectiveness of economic activity and social policy, therefore ultimately driving improved economic and social outcomes.

1.3. Assessing the opportunity and benefits

Global scale

In a recent report, McKinsey has estimated that US\$5 trillion of economic value can be unlocked by enabling greater liquidity of data¹. Clearly this is a huge economic opportunity.



¹ <http://www.mckinsey.com/business-functions/business-technology/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>

Likely evolution of data sharing on a global scale

It is important to consider how the underlying industry that *facilitates* safe data exchange will evolve globally. Major global businesses are emerging that dominate their respective marketplaces or exchanges. This is currently the case in ride-sharing, holiday accommodation, hotel, flight & travel aggregation, retailing, online auctions, jobs and social media. We believe that this is likely to be the case with data exchange where one or a few bodies will dominate globally.

Data exchange is likely to take place initially in a large number of hubs that facilitate exchanges between groups of organisations, potentially clustered within verticals. We expect that this will consolidate rapidly into a smaller number with just one or a few leading exchanges, as has happened in other content driven platforms. The breadth of data exchanges that each hub can facilitate will determine their individual success. Those offering the greatest utility will prevail.

The opportunity for Australia

Governments around the world are seeking to enable data sharing via a variety of different approaches. Most are focussed on the benefit of unlocking data however, none have focussed on fostering the underlying industry of data exchange with a view to becoming a global centre of excellence in the same way that, for example, London or New York are centres for financial services.

It is our view there is a clear opportunity for Australia to be a global leader in the data sharing industry. Australia has a well-developed legal and privacy framework with advanced IT industries and trusted political and governance institutions. Many components required for this eco-system are already in place and growth could easily be fostered with appropriate government support and policy.

The benefits of Australia leading this industry are:

- Unlock enormous productivity improvements nationally,
- Generate significant employment opportunities in high-technology industries,
- Retain intellectual property in Australia,
- Position Australia as a leading destination for data scientists,
- Boost the standing and capability of universities as centres of research,
- Provide national security benefits.

We recommend that the Privacy Commission considers how Australia could become a leader in this emerging industry and how government could use policy to foster growth.

1.4. Market need for data sharing

The Productivity Commission's investigation into data sharing is timely. Of the 250 organisations Data Republic have spoken to, almost all are seeking a way of enabling data sharing but most are finding it to be extremely difficult, time-consuming and expensive. Often the barriers are so great that data sharing initiatives simply fail.

Data has been exchanged between organisations for a long time. However, it is clear from our conversations with organisations that this is often done in an ad hoc way, with few, if any, governance practices. This creates risks for both organisations and the people whose data the organisations hold.

2. Introduction to Data Republic

2.1. Our business objective

Data Republic was established to enable organisations to share data in a safe, secure, privacy-compliant and well-governed manner. We believe that the benefits of data sharing, when conducted safely, are enormous and will generate major economic and social benefits.

Information held by organisations is becoming more liquid and increasing in velocity as it flows around the economy. However, it lacks a solid framework for enablement that will provide transparency to the organisations and to consumers.

2.2. Our experience over the last two years

In the two years up to August 2016, when this submission was prepared, Data Republic's team has spoken about the concept of data sharing with over 250 organisations in Australia, plus many more overseas. These organisations have included private sector companies, government departments and not-for-profit organisations. Of the 250 organisations:

- 80 are ASX200 listed companies,
- 50 are multi-nationals headquartered overseas,
- 30 are government departments,
- 20 are not-for-profit organisations,
- 70 are non-ASX200 Australian companies.

This submission has been prepared to reflect our findings from these detailed conversations. We believe we are well placed to outline the issues relating to data sharing that are of concern to organisations, and the solutions that may be appropriate.

3. Risks and issues related to data sharing

The following sections cover the risks and issues that we have uncovered in our conversations with 250 organisations in the last two years.

The combination of these issues means that data sharing is very hard, expensive and time consuming to do. Many organisations have tried and failed to establish exchanges of data, generally being defeated by cost, time or lack of risk mitigants.

3.1. Governance

Issues

Governance is the most significant issue that we have encountered, especially amongst senior management who carry ultimate responsibility for control of data. When releasing data, organisations want to be able to audit at all times:

- Exactly what data is leaving the organisation,
- Which organisations are receiving the data,
- Exactly how the data are being used,
- Who has access to it,

Establishing this knowledge for each data exchange is time consuming and hence results in delays.

Solutions

The framework for requesting and approving exchanges of data must be flexible enough to control every aspect of the exchange process. Data Republic's trust framework was designed to enable every cell of data to be controlled, and for the permitted use and end user to be specified, approved and monitored. In turn, this enables organisations to have a complete view of where and how their data is being deployed.

This is a crucial point for governments who are increasingly looking towards open data policies. In our review of many government 'open data initiatives' around the world, all have a 'black and white' approach where individual datasets are either completely shared, or not shared at all.

This precludes the various shades of grey in between that might lead to significant beneficial policy outcomes. For example, there will be many government datasets on health and crime that could not be released on an open data portal but which could be very valuable in the hands of a university researcher or policy analyst. Good governance will enable selective sharing of data.

3.2. Privacy guidelines adherence

Issues

The Australian Privacy Principles Guidelines (APP) require interpretation by specialists in privacy law. This interpretation process often requires review before, during and after solutions have been implemented to ensure that the process meets expectations from the APP.

The impact of this is significant. It is not unusual for data exchanges to take up to six to twelve months, with the majority of time spent on legal and IT issues.

Solutions

Accelerating data sharing as an activity requires a fast-track method for making all parties comfortable with the way that data is being exchanged

Data Republic developed a legal framework and business model that was ‘private by design’ from the outset so that participants could feel comfortable that APP was being adhered to. This is a single agreement that all participants enter into, which creates transparency and openness. The agreement also enforces that organisations that provide data have the right to do so, including consent from consumers.

3.3. Risk of identification of personal information

Issues

The linking of personal information from two sources is a crucial part of the highest value data exchanges, even if the output of those exchanges is not related to individuals. However, releasing personal information with attributes of those people is too high risk for many organisations and so exchanges do not take place.

There is, however, a clear difference between the “inputs and analysis” of data and the “outcomes of analysis”. For example, it might be valuable to link data on health outcomes to data on long-term diet from grocery shopping behaviour. The matching of these two datasets is difficult to achieve without high risks, but the outcome of the analysis itself is likely to be innocuous and low risk e.g. if it is used to target diabetes education to segments of people in particular areas.

Solutions

There are several ways that this can be facilitated. However, Data Republic has separated and anonymised personal information (name, address, email etc.) from attribute data (has a loyalty card, buys baked beans, drives a Toyota). Personal information is held in Data Bank, located within Westpac’s firewall. Attribute data is held in Data Republic. The two types of data are linked in a confidential and secure way.

3.4. Legal negotiations

Issues

This issue is related to the complexity of interpreting APP and applying the guidelines to technical solutions, and identifying that the risks and controls are commensurate with the organisation’s risk appetite. With data exchanges being high-risk activities, legal teams are often compelled to create highly engineered contracts that cover every eventuality and reduce risk towards zero. We have seen that the negotiation of these contracts can take many months and so make data exchange inefficient or prohibitively expensive.

Solutions

Data exchange requires a common agreement that enables multiple organisations to exchange data with each other under standard terms. The concept of ‘joining a club’ is understood and accepted by Data Republic’s participants in order to create transparency and openness. Ultimately all data exchange contracts will require the same issues to be dealt with.

3.5. Time and cost to implement

Issues

Both commercial and government organisations mention that the negotiation of data exchange or data access with a second party could extend over 6 months in commerce, or 12 months in government. Legal and other resources are applied, both adding to the elapsed time and escalating costs.

Solutions

As outlined above, common contracts and IT frameworks that are private by design can rapidly accelerate data exchanges from months to days.

3.6. Risk of re-identification of personal information

Issues

If a large number of characteristics about an anonymised individual are present, there exists a risk of re-identification. This occurs when the combination of characteristics is unique. For example, there is only one 42-year-old male living in Randwick NSW, with two children aged 4 and 7 that owns a Kia Sorrento.

Solutions

Ensuring that any visible data does not display unique records, whilst still allowing meaningful analysis, can eliminate this risk. We have worked with Data61, part of CSIRO to utilise technology called 'differential privacy' that performs this function.

3.7. Security

Issues

Security of data is fundamental to data sharing and data movement. IT security is a highly specialist topic, and the subject of an ongoing arms race between system defenders and system attackers. Many CIOs that we have spoken to naturally have a concern that once data leaves their own systems, they have effectively lost control of it whilst retaining accountability for the confidential storage of data

Solutions

The requirements for secure exchange of data are manifold. Based on the conversations we have had, with organisations themselves and also with experts in best practice, it is increasingly apparent that the following are essential

- High security storage environments, with the cloud now providing the safest harbour.
- Separation of personal information from attributes about those individuals. We elected to store personal information behind bank grade security, behind a bank's secure firewall, creating essentially a 'bank for data'.
- Data is stored encrypted at rest
- Data from multiple data owners is stored separately
- Users of data can only access with permission from the data owner via a process of request and approve on platform, resulting in a discrete audit trail.
- Analysis can be facilitated in virtual workspaces, whereby control and oversight of data is maintained. This enables analysis to be done on sensitive data, without the risk of it being inadvertently leaked.

3.8. Safe, neutral harbour for data

Issues

Organisations are overwhelmingly more comfortable in *receiving* data than *giving* data in a data exchange. Giving data to other organisations reduces the control that the giver has, and contributes to risk. This effectively prevents many data exchanges from taking place as neither side can find enough comfort to release data to one another.

Solutions

Data Republic elected to provide a Safe Harbour for data in the cloud. With appropriate security, the one-way nature of data exchanges is eliminated and both sides feel that they are involved on an equal footing.

Furthermore, the analysis of data is conducted in a controlled environment where data can be seen and analysed but not downloaded without consent. This enables sensitive analysis to be conducted in a secure environment where data contributors know that their data are protected.

3.9. Pricing and valuing data

Issues

Many companies are now seeking to recognise the value of their data assets, both as a form of new business revenue, and as a potential asset to be represented on their balance sheets.

There is currently no mechanism to independently assess the value of data, both to set the price for sales or value the asset.

Solutions

Data Republic believes that a consolidated data exchange will eventually have enough metadata and insight on the value of data to be able to independently set pricing and provide a valuation model for data. This will not be as achievable with multiple, fragmented exchange hubs.

3.10. Lineage

Issues

As data become more liquid and moves at a higher velocity, users of data will have less, if any, direct contact with the owner of the data. This will raise concerns about quality and authenticity. The risk is that poor quality or illegally used data are in market, causing similar problems to having fake banknotes in circulation.

Solutions

Ensuring that the lineage of data is documented and certified, remains a crucial part of enabling data exchange. We believe that a strong legal and governance framework is best placed to facilitate this.

3.11. National security and fraud

Issues

The fragmentation of data provides loopholes for criminals. Money laundering is one such example where government has drawn data from Australian banks to investigate money laundering. Nevertheless, there are benefits to further exchange of data both domestically and with international co-operation.

There are a number of domestic anti-fraud applications that could be enabled by safe data exchange. The finance and insurance industries are examples that could benefit from greater co-operation to prevent fraud, but without the risk of releasing commercially sensitive information.

Solutions

Data exchange mechanisms must provide the ability for private collaboration.

3.12. Giving consumers control of their data

Issues

It is essential that consumers have an easy way to understand how their data is being used, and prevent it from being used if they do not personally approve it.

Solutions

We believe that transparency and openness is essential for data exchange to become a mainstream economic activity. A single, consolidated store of personal information will give consumers the ability to opt out of all, or just specific types of data share. A fragmented approach to data exchange will not permit this.

In our view, it is unlikely that this will occur by individual consumers working together to ask companies for greater control, but will be facilitated by companies safely pooling all of their data and then enabling consumers to opt out.

Data Bank is the vehicle to store personal information within the Data Republic ecosystem. We foresee that consumers will eventually trust banks with their personal information in the same way that they currently trust banks to safely hold their money.

3.13. Data exchange is only possible for large organisations

Issues

As noted above, the IT, legal and senior management resources required to facilitate safe data exchange are considerable. Data exchange is therefore something that is generally only practiced by large companies that possess the necessary resources. To unlock the full economic value, and to facilitate innovation, data exchange has to be made available to a broader range of companies, government bodies and not for profits.

Solutions

Data Republic has created a centralised platform for data exchange that can be accessed by any organisation, no matter how small. We have invested in the legal and IT infrastructure to enable the democratisation of data exchange.

We have also set up a not for profit body, The Minerva Foundation, to enable data contributors and data scientists to provide pro bono assistance to not for profits to see

data your for positive social outcomes. We have already completed a number of assignments for charities.

3.14. Data Sovereignty

Issues

Data sovereignty is becoming a lesser concern as cloud becomes more widely accepted. However, many organisations now use offshore agencies for data analytics, which is perceived to reduce the amount and quality of data governance. Data relating to the *attributes* of individuals should not pose a great concern, but *personal information* is a greater risk.

Solutions

Data Republic elected to use Amazon Web Services' Sydney regional cloud infrastructure. This has removed the issue of data sovereignty for attribute data

Data Bank is a place for personal information to be 'banked' – held securely in the same way that banks are trusted to hold money. The Australian instance of Data Bank is within Westpac's bank-grade security. We believe that a separate instance of Data Bank will be required for each nation, and we are in discussions with banks in the United States, the United Kingdom, Europe and Asia about opening local Data Banks.

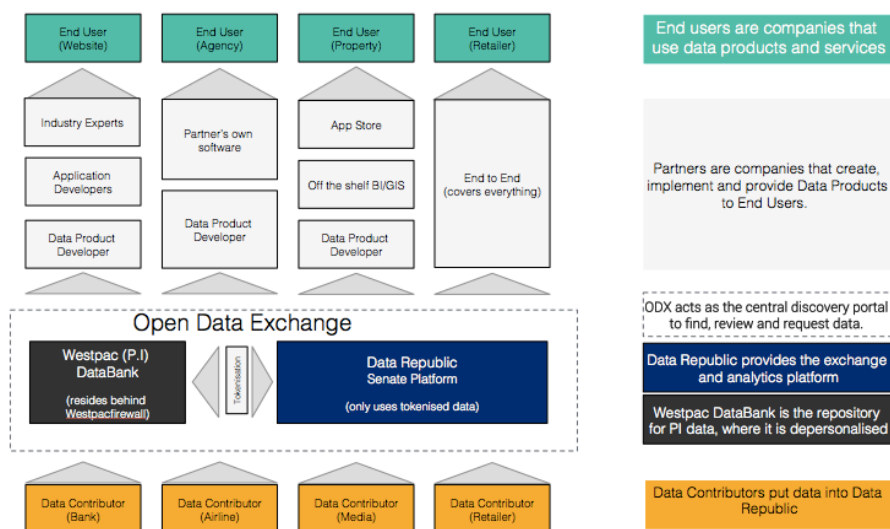
4. Background on Data Republic

4.1. Company overview

Data Republic was established in 2014 in Sydney. The company now employs 27 people in Australia. Our United States office opened in 2016 and now employs two people. Data Republic enables companies to exchange data in a safe, secure, privacy-compliant and well-governed manner.

4.2. Data Republic's ecosystem

The diagram below outlines the ways that the flow of data is facilitated.



4.3. The future

We are currently exploring further overseas expansion. We have received expressions of interest from companies and governments in Europe and Asia to facilitate data exchange.

4.4. Company structure

Data Republic is wholly owned by its staff and five major investors:

- Paul McCarney as co-founder and initial investor,
- Reinventure, Westpac's venture capital fund,
- NAB Ventures, NAB's venture capital fund,
- Red Planet, a division of Qantas,
- Qualgro, a venture capital fund based in Asia.