All Your Data Are Belong To Us – Trends in Data Ownership

Prepared by Chris Dolman and Hugh Miller

Presented to the Actuaries Institute
20/20 All-Actuaries *Virtual* Summit
3-28 August 2020

Abstract

"Data is the new oil" – use of data is central to modern business success and it is no surprise that all four non-state companies to achieve trillion-dollar valuations are technology companies that rely heavily on data. But how can data be accessed? Who owns data? What control do you have over your data? What happens when people do not have enough data assets, like credit or banking history, to access services that increasingly require it? This paper considers trends in data ownership and some of the implications of these trends, to help actuaries stay up to date in this fast-moving area.

First, we look at the implications of the current Australian Privacy Principles. Although they give significant rights to individuals, the legal interpretation of what constitutes personal data highlights ongoing challenges. Full compliance for an access request typically requires sophisticated data management systems and constant vigilance. We look at personal examples of data requests made under the APPs, to illustrate the challenges which can be faced. We will also discuss practices companies are using to comply with privacy rules, including transforming personal data to detailed 'non-personal' data. We explore what businesses should consider doing to meet the expectations of customers which may go beyond strict compliance with legal obligations.

Second, we explore trends in data portability including recent open banking rules. The intended operation of the rules is to give people greater ability to access, use and redirect data about themselves that is currently held by institutions. This is predicted to enhance competition and innovation. However, when applied as a broader theme across many sectors, there could be significant side effects, particularly around product access for those who are digitally disadvantaged. Balancing the benefits with the side effects will be a significant challenge as these systems mature.

Finally, we discuss recent developments on the data regulation front and the implications for further developments that may significantly affect how companies collect and use data.

Keywords: data; privacy; regulation; access; portability

Who "Owns" Data, Anyway?

To explore the ownership of data, we must first think about what data actually is. In common parlance, data typically refers to a digitised observation of information relating to a person, object or some form of interaction. For our purposes here this common (albeit vague) definition is sufficient - though we note considerable academic work on information theory separating the concept of data from information, this subtlety tends to be lost in popular policy discussion of "data ownership".

So, if data is defined as above, it cannot be "owned" in the same way as a physical object like a house or car, nor innately possessed like your body. Data is virtual, not physical, in which case, "ownership" could mean some exclusive right or control over this virtual property - similar to the concept of intellectual property. However, this need for exclusivity is a problem for data "ownership", since:

- Data can be created which two or more parties might rightly claim "ownership" of (for example data recording a bank transaction between them). How, then, should exclusivity be managed? May one party to the transaction restrict what the other party may do with the information?
- Data can be created about an object, without controversy, but this data may in future be associated with a person who might then wish to claim ownership of that data. The service history of a car is not personal data, but perhaps if you also know it's *my* car, this tells you something personal about me (for example whether I'm the sort of person to be diligent about car servicing).
- Data can be created and replicated at near-zero cost, by anyone with access to the data, so it may be difficult for a person to know who exactly has a copy of a particular piece of data about them. This can make it difficult to exert any powers of ownership that we might create.
- Data may be created by someone without your knowledge. For example, a CRM system may record comments from sales staff about a prospect, without the knowledge of that prospect. Again, it is difficult to exert a power of ownership over something you have no knowledge of.

Notwithstanding the difficulties listed above, there is an increasing trend of data ownership rights around the world. This is, in effect, the creation of incremental rights of control over how certain data relating to an individual may be created, stored and used.

What Rights do you Have Over Your Data, Today?

Australia has relatively clear data ownership regulations. Much of this is tied to privacy law which specifies what people are allowed to do with data. The Privacy Act 1988 is the main governing legislation and it enshrines the 13 Australian Privacy Principles (APPs), which is focused on personal information (information about a reasonably identifiable individual). The Office of the Australian Information Commissioner (OAIC) regulates privacy in Australia and has good coverage of the APPs. We very briefly summarise these APPs in the table below. We refer to 'organisations' in a loose term here - we generally mean entities that are covered by the APPs. There are also exceptions to almost all principles.

Table 1: Summary of the Australian Privacy Principles

APP	Informal summary
1: Open and transparent management of personal information	Organisations must have a privacy policy that transparently describes how personal information is managed
2: Anonymity and pseudonymity	Unless authorised by law or is impractical, individuals must have an option to remain anonymous.
3: Collection of solicited personal information	Organisations can only collect information that is reasonably necessary for its activities. There are higher standards again for sensitive information. Personal information should be collected from that individual.
4: Dealing with unsolicited personal information	If an organisation receives personal information that fails APP3, it should delete it
5: Notification of the collection of personal information	Organisations must let people know they are collecting personal information, and describe the purpose of collecting the information
6: Use or disclosure of personal information	Organisations should not use or disclose personal information for a secondary purpose unless consent is granted, or a person would reasonably expect it, or required by law.
7: Direct marketing	Similar to APP6, an organisation must obtain appropriate permission for the use of personal information for direct marketing, and must also give an opt-out mechanism
8: Cross-border disclosure of personal information	For overseas disclosures, must take reasonable steps to ensure the overseas entity complies with APPs
9: Adoption, use or disclosure of government related identifiers	Organisations should not use government identifiers
10: Quality of personal information	Organisations should take reasonable steps to ensure personal information is accurate, up-to-date and complete.
11: Security of personal information	Personal information must be kept secure and deleted or de-identify the data if it's no longer needed.
12: Access to personal information	Organisations must give access to personal information about an individual on request, in a reasonably timely manner
13: Correction of personal information	If an individual says personal information is wrong, the organisation needs to correct it.

We make some general observations:

- APP2 feels poorly reflected in society; very few organisations make it easy to create queries anonymously. My bank requires detailed identity checks before answering relatively general questions.
- APP3 is being stretched and perhaps the general intent is being challenged in a digitized environment. Is it "reasonably necessary" for a supermarket to know every item I've ever bought to offer me service?
- APP11 requires deletion or de-identification when data is no longer needed; we believe this runs contrary to default practices for some organisations, although we know some are quite disciplined around data deletion.
- Disclosure is a prominent feature across the APPs, but we know that one consequence of this is overly long privacy policies and other terms and

conditions documents that nobody reads. But they can make good art – see **Figure 1**.

Figure 1: Dima Yarovinsky prints out the 'terms of service' of leading online services for display at the Bezalel Academy of Arts and Design

How do Australian rights compare internationally?

We offer some brief comments comparing Australia to Europe and the USA:

- The EU General Data Protection Regulation (GDPR) is a recent regulation that provides some of the strongest privacy rules internationally. It similarly focuses on personal data, including special categories.
    - It has expanded accountability and governance.
    - It is quite clear that 'consent' must be freely given, specific and informed.
    - It includes a 'right to be forgotten' that an individual can request
    - Fines for non-compliance are potentially very large (up to €20 million or 4 per cent of annual worldwide turnover)
- The United States does not have overarching legislation, but a series of laws around specific areas (health, financial services), electronic communications and children. In recent years regulation has reduced in certain areas rather than increased. For example, collection of browsing history required consent prior to 2017.

So, Australia appears 'somewhere in the middle' on overall protections but benefits from having dedicated laws and a regulator to examine privacy issues. Critics of the Australian setup point to the lack of redress (there is no right for individuals to sue for privacy breaches), and the lack of legal standing overseas (particularly in the USA).

What actually is personal information?

The current Privacy Act definition of personal information is:

'Information or an opinion about an identified individual, or an individual who is reasonably identifiable:

- whether the information or opinion is true or not; and
- whether the information or opinion is recorded in a material form or not.'

The guidance on the OAIC website[1] gives a long series of examples describing its interpretation of this. Separate datasets where the person's name is on one, but a common reference number is on both, is viewed as reasonably identifiable (but the second table, if passed to a consultant, may not be). Information may relate to more than one person, or may relate to a person and something else (e.g. body corporate fee payments may relate to a person but also a property). It could relate to data beyond records such as voice recordings and DNA.

This appears broad, offering good protection for consumers. However, the only significant court case testing the definition that the authors are aware of has created significant controversy. The case was *Privacy Commissioner v Telstra Corporation Limited*, determined in the Federal Court in 2017. The case concerned Ben Grubb, an ABC journalist, seeking to access his personal information from Telstra. Telstra provided some data, but withheld geolocation data, claiming that this was anonymous.

The Privacy Commissioner agreed and ordered Telstra to hand over the data. Telstra appealed and the Administrative Appeals Tribunal (AAT) found in favour of Telstra. They found that the location data was 'about a mobile device' rather than 'about an individual'.

A key in the judgement analogy was car maintenance; service information ('log book') relates to the car, rather than an individual, even if the registration number on the service book can be tied to the owner via another record.

The Privacy Commission appealed to the Federal Court, but only on a fairly narrow topic, and lost.

The case raises significant questions of whether information should be deemed personal, even if it can be linked back to an individual. Taken at face value, it raises the possibility that personal information can be defined by organisations more narrowly than other readings of the law, and perhaps more narrowly than a layperson might reasonably expect.

We note that the government's anti-terror metadata retention program - under the *Data Retention Act*, slightly modifies the definition of personal information and makes some explicit statements around retained metadata being personal information. However, the broader case findings that information may not be 'about an individual' may still apply in other contexts. For instance:

- Is shopping transaction data about an individual or about a set of bought items?
- Is information held for home or CTP insurance about an individual or about a home and car?
- Is internet search history about an individual or about a web browser or device?

The issue has not been rectified through updated legislation. The European rules, GDPR, defines personal information as *any information relating to an identified or identifiable natural person*. This definition is harder to interpret narrowly, as 'relating to' is broader than 'about'.

Notwithstanding the issues above which could limit the legal rights offered to individuals seeking access to data about them, we feel that it is likely that in some contexts, some consumers might expect a broader rather than a narrower interpretation of this definition. Businesses should bear this in mind in any actions they take. Conduct which is perhaps technically legal but still a breach of public expectations is still a problem for a business and should be taken seriously. At the very least such conduct will lead to dissatisfied customers, which no business wants. We suggest that if customers are seeking access to their personal data and their request might go beyond strict legal requirements, it may still be in an organisation's best interests to provide this data. Clearly there will need to be an assessment of both practicality and cost when considering this.

Anecdotally, we are aware that some businesses are very proactive in protecting personal information. This can be supported by policies such as periodic deletion of older records, and strict data silos that cannot be used for marketing or third-party purposes. Some companies with exposure to Europe have taken a 'international best practice' approach and applied globally the requirements of GDPR, which includes robust handling of personal data.

How Easy is it To Access Data About You?

There is a practical way to test how well set up companies are to comply with Australian Privacy Principles, particularly APP12 (access to data) – we can ask them for our data. One of the authors (Hugh) did that for a large number of companies in early 2020. These companies are listed in the table below.

The personal data requests were generally made online (via email where possible, online forms otherwise). For the four technology companies at the end of the list we used the self-service data facility. The main round of requests was done in mid-March, with sporadic follow-up thereafter. The requests contained:

- A reminder of an organisation's responsibilities under the APPs
- Personal information held
- List of any third parties data was provided to
- Some detailed items (e.g. transaction data, location data), tailored depending on the organisation.

Table 2: Summary of personal data requests made

Company	Summary	Customer Service	Data provision
Internet provider (TPG)	• Initial referral to correct privacy address (not available on website) • No response - 3 months.	✗	
Mobile provider (Vodafone)	• Reasonably fast follow-up • Only limited metadata available (data sessions, call times), at a fee.	✓✓	✓
Home insurance (Allianz)	• Responsive follow-up • Letter addressing some of the specific queries raised (e.g. identified 6 marketing campaigns that I had been targeted, via a NZ distributor) • Actual personal data provided as a series of screenshots from policy management system.	✓✓	✓
Life insurance (AIA)	• No response.	✗	
CTP insurance (QBE)	• Timely response from the 'complexctp' team • Email described all disclosure. • Nothing held beyond the minimum for pricing and maintaining the policy .	✓✓	✓✓
Health insurance (HIF)	• No response.	✗	
Bank (St George)	• No response until I followed up later as a complaint.	✗	

Supermarket (Woolworths)	• Additional paperwork for request supplied, indicating a process
	• Multiple files provided (password protected), including
	  ○ Every item bought since 2014 attached to loyalty card. 113 page pdf
	  ○ 26 pages of campaign lists
	  ○ 6 pages of clicks for campaigns.
		✓✓	✓✓

Data broker - Experian	• No response.	✗	
Data broker - Equifax	• No response.	✗	
Data broker - Acxiom	• No response.	✗	
Airline (Qantas)	• Timely response, only flight history and point earning history. Nothing on website use or marketing offers.	✓✓	✓✓
Airline (Virgin)	Perhaps the most detailed • Member profile and comm preferences • Flight bookings • Paid media campaigns • Marketing campaigns (emails) • Paid marketing campaigns (to ad networks) - 63 over three years.	✓✓	✓✓✓
Face recognition (Clearview AI)	• Required personal details before they proceed • Did not respond until I followed up months later. Then a fast turnaround.	✓	✓

Company	Summary	Customer Service	Data provision
Tech (Facebook)	Comprehensive self-service tool. Includes lots of potential advertising and third-party information.	✓✓	✓✓
Tech (Google)	Comprehensive self-service tool.	✓✓	✓✓
Tech (Microsoft)	Comprehensive self-service tool.	✓✓	✓✓
Tech (Amazon)	Purchase history, but very little else available.	✓✓	✓

We make some observations from this anecdotal experience.

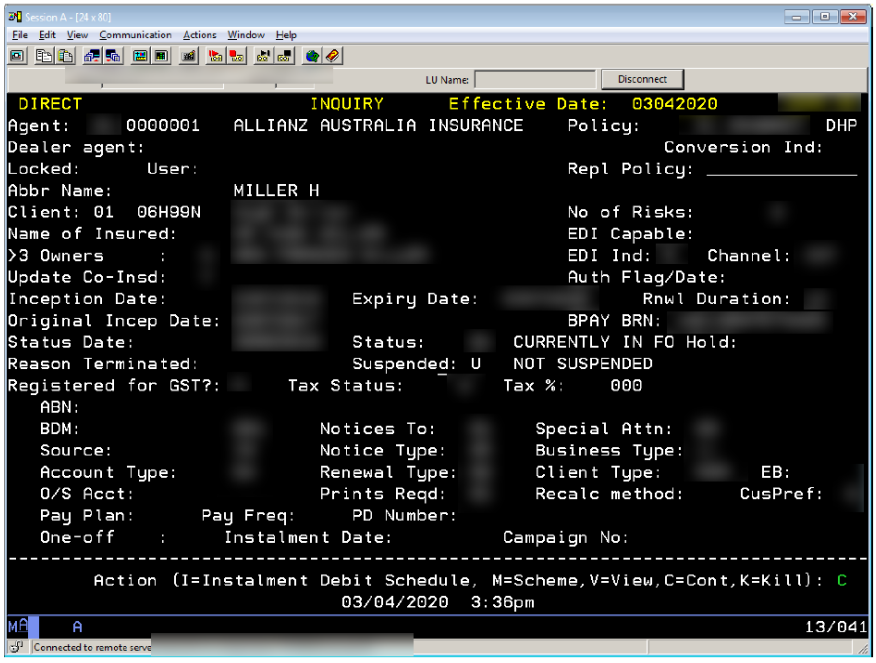
There were big differences in maturity for handling this request

Most organisations had a dedicated email address for privacy-related inquiries. Some companies seem hard to contact - the technology companies being the most obvious, while other companies prefer online forms to email addresses.

The technology companies are definitely in the lead regarding access to personal data. Facebook, Google, Microsoft and (to a lesser extent) Amazon all had well-established self-service options for collating and downloading data, much of it very comprehensive. Of course, these are also the companies who capture the most data about us, so perhaps it makes sense that their processes are more mature.

In contrast, some companies appear to have very manual processes. The home insurer provided a series of screenshots from their policy management system. Perhaps this approach is understandable when volumes of enquiries are low, as might be the case in this example.

Figure 2: Home insurer provision of personal information



Some of the non-tech companies appeared to have mature systems. Vodafone, for example, appeared to have a well-developed policy and team.

Based on this exercise, not everyone appears compliant with APPs

Some companies did not respond to the inquiry. In theory this could make them in breach of their obligations under APPs, but we note two things:

1. I didn't try following up very hard
2. The bulk of the first enquiries were made mid-March 2020; this turned out to be a busy period for many companies dealing with a pandemic and lockdowns

Further, the amount of details provided by some companies seems low relative to what we would have expected. As shown in the earlier discussion, there is potential uncertainty about what falls under the scope of personal data, but we suspect that sometimes organisations provide what is easy rather than systematically working through all their data. It's unclear how much more data might have been provided, had we pushed for it.

I have no idea what is going on with data brokers

Data brokers like Experian, Equifax and Acxiom claim to have large customer databases with collected characteristics. None of them responded to the enquiries, and, at least as it relates to me, it is unclear exactly what they hold and how much of it is deidentified versus personal.

Most companies are not doing anything unexpected with my data

Giving personal data to third parties was common to all data requests. Most did not identify any such disclosures and many companies were explicit on their policy not to disclose.

Allianz acknowledged a third party for sending out their marketing emails.

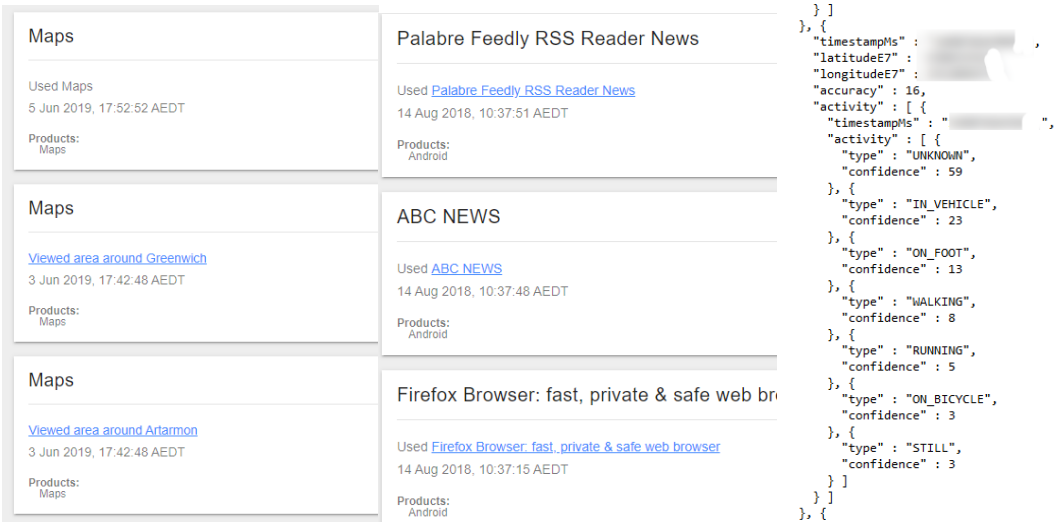
Some companies, particularly technology companies, collect a lot

We know this, but there were still a few surprises in the requests. Most have good options for turning off various forms of tracking. It turns out for around 1 year I had activity tracking active with Google, which led to items such as:

- 12,000 app opening items
- A bunch of audio, not always my voice
- A full history of locations examined on google maps.

Some example items are shown in **Figure 3**.

Figure 3: Example Google history for map searches, app use and location history



Woolworths provided me with 113 pages of information, showing every item for every shop. Depressingly, my typical shopping basket hasn't changed very much over five years.

There is a lot of potential information missing

The difference in airlines was interesting; Qantas only provided booking and points, whereas Virgin provided extensive lists of campaigns (both direct marketing and cases when I was included in web advertising campaigns). Virgin sent me 338 direct marketing campaigns over 5 years, of which I only clicked 3 times; perhaps not their best investment?

Virgin also put in marketing segment information (extract included below). I'm generally a good (but not great) prospect in a lot of their propensity models. This type of analysis probably goes on across many companies, but rarely was provided in the data requests.

Figure 4: Virgin propensity scores

```
AMEXPLAT_CC_PROPENSITY_DECILE="2"
CAR_EARN_PROPENSITY_DECILE="2"
DECC_EARN_PROPENSITY_DECILE="3"
FLYING_BEHAVIOUR_RULE="H"
HOTEL_EARN_PROPENSITY_DECILE="3"
IECC_EARN_PROPENSITY_DECILE="3"
INT_AIR_EARN_PROPENSITY_DECILE="2"
NEXT_BEST_REWARD="empty"
REDEMPTION_SEGMENT="6"
INT_SHARESHIFT_TARGETCELL="Target"
CC_ANZ_PROPENSITY_DECILE="3"
BP_PROPENSITY_DECILE="4"
FLYBUYS_TRANSFER_DECILE="5"
VMA_CC_PROPENSITY_DECILE="3"
VMA_HL_PROPENSITY_DECILE="1"
EBY_POINTS_BAND="e7800-15599"
EBY_POINTS_BAND2="c4000-11799"
CC_AMEX_ESC_PROPENSITY_DECILE="5"
WBC_HL_PROPENSITY_DECILE="1"
```

Another interesting area is the interpretability of data provided. For example, supermarket campaign information is just a series of codes; the information about the type and nature of that campaign is potentially still personal information but was not provided.

Figure 5: Supermarket campaign listings

Date and Time	[redacted]	Campaign Code	Campaign ID
2020-03-20 03:01:10		ADH-3258	10717
2020-03-15 20:12:17		ADH-3267	11016
2020-03-13 05:42:00		ADH-3258	10717
2020-03-10 19:13:59		CVM-1661	37
2020-03-09 00:18:27		CNA-0005	500
2020-03-05 02:30:44		ADH-3258	10717
2020-03-03 14:46:00		CVM-1661	37

Additionally, it is unclear whether information on me collected when I do not scan my loyalty card has been included. As noted in the 2019 ACCC Loyalty Card report, both Woolworths and Coles are allowed under their privacy policy to track linked transactions even when we don't scan. It is unclear whether these have been included.

For non-record data, we note that no organisations offered recorded phone calls, and tech companies did not offer tagged photos or videos. These potentially fall within the scope of personal data.

Most datasets appear correct

We've not done a thorough audit, but computerisation meant that the chances of errors are limited. Two issues that did stand out:

- Facebook includes a list of organisations that have uploaded my details so they can target me within Facebook. This list was very long - over a thousand; it was dominated by US-specific companies (mainly car sellers). Dates were not provided. The nature of the list suggests that my details have either snuck onto a random data broker list in the USA, or I have been fuzzy matched poorly. The list included Woolworths, which is likely legitimate.
- Woolworths purchase data included a swathe of duplicates over a period of time; it is unclear if the duplication is in the underlying data, or whether it was an error introduced while pulling together the response to the personal data request.

Figure 6: Examples of companies who uploaded a contact list with my information

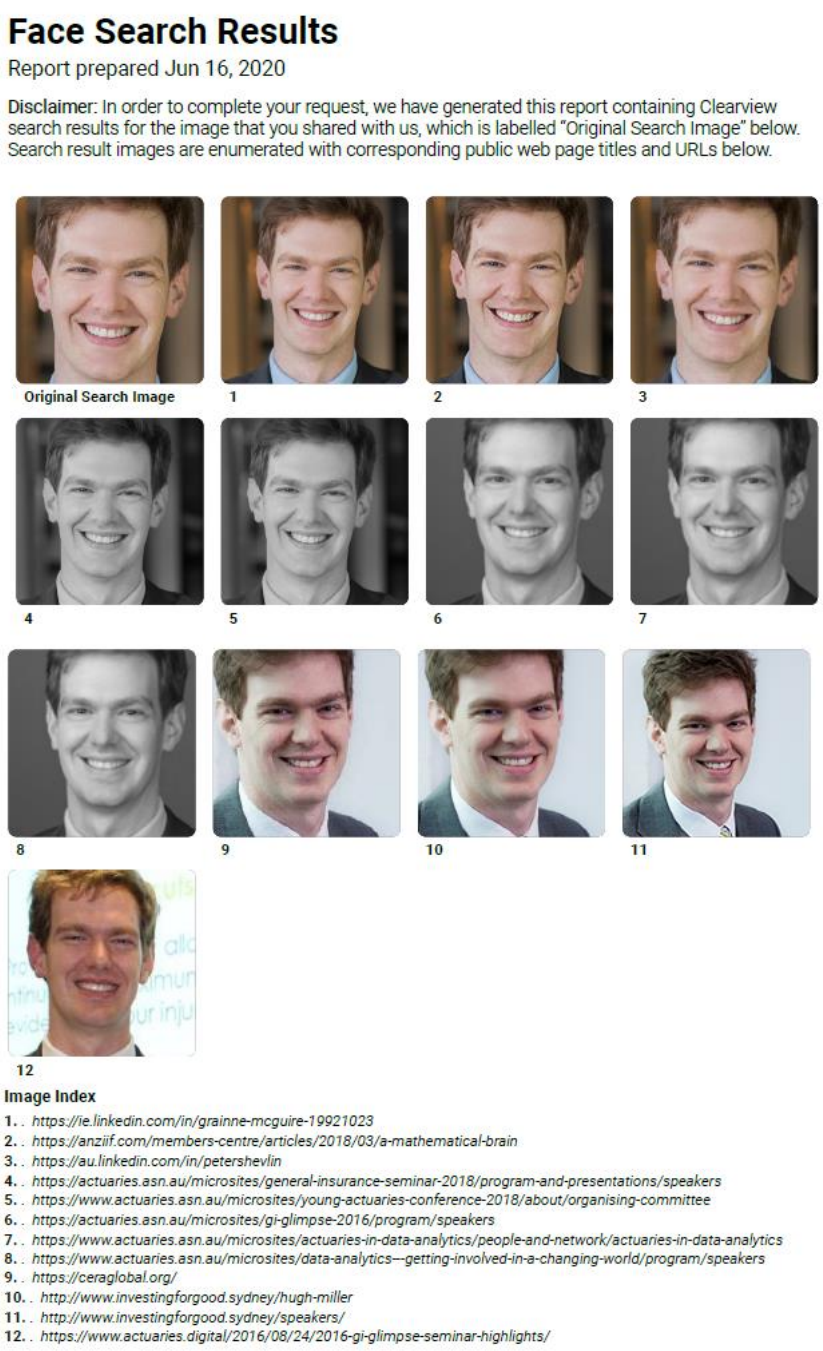
Figure 7: Duplicate information in Supermarket records

Mandarin Afourer Md	1	2.66
Mandarin Afourer Md	1	2.66
WW Oats Quick 750g	1	1.09
WW Oats Quick 750g	1	1.09
Traditional Fruited Bun 6pk	1	3.48
Traditional Fruited Bun 6pk	1	3.48
WW Cheese Slices Swiss Light 200g	1	4.48
WW Cheese Slices Swiss Light 200g	1	4.48
Chobani Limited Batch Flavour 170g	2	2.98
Chobani Limited Batch Flavour 170g	2	2.98
Zucchini Green Kg Md	1	0.65
Zucchini Green Kg Md	1	0.65
WW Homestyle Burger 500g	1	5.47
WW Homestyle Burger 500g	1	5.47
WW Cheese Shredded Mozzarella 500g	1	4.48
WW Cheese Shredded Mozzarella 500g	1	4.48
WW Cheese Slices Tasty Light 500g	1	5.97
WW Cheese Slices Tasty Light 500g	1	5.97
Wonder Smooth Wholegrain 700g	1	3.18
Wonder Smooth Wholegrain 700g	1	3.18

Facial recognition technology has significant privacy implications

Clearview AI is a relatively new technology start-up that has scraped billions of face images from the internet and turned it into a reverse image search tool for faces, allowing an organisation (whether a police force, retailer or other) to find other images of that person to help identify them. It has been used (or at least piloted) in Australia[2] by police and has attracted controversy, particularly in the USA. It certainly has Hugh on file.

Figure 8 Clearview AI has more photos of Hugh Miller than you will ever need



Face Search Results

Report prepared Jun 16, 2020

Disclaimer: In order to complete your request, we have generated this report containing Clearview search results for the image that you shared with us, which is labelled "Original Search Image" below. Search result images are enumerated with corresponding public web page titles and URLs below.

Original Search Image 1 2 3

4 5 6 7

8 9 10 11

12

Image Index

1. . https://ie.linkedin.com/in/grainne-mcguire-19921023
2. . https://anziif.com/members-centre/articles/2018/03/a-mathematical-brain
3. . https://au.linkedin.com/in/petershevlin
4. . https://actuaries.asn.au/microsites/general-insurance-seminar-2018/program-and-presentations/speakers
5. . https://www.actuaries.asn.au/microsites/young-actuaries-conference-2018/about/organising-committee
6. . https://actuaries.asn.au/microsites/gi-glimpse-2016/program/speakers
7. . https://www.actuaries.asn.au/microsites/actuaries-in-data-analytics/people-and-network/actuaries-in-data-analytics
8. . https://www.actuaries.asn.au/microsites/data-analytics---getting-involved-in-a-changing-world/program/speakers
9. . https://ceraglobal.org/
10. . http://www.investingforgood.sydney/hugh-miller
11. . http://www.investingforgood.sydney/speakers/
12. . https://www.actuaries.digital/2016/08/24/2016-gi-glimpse-seminar-highlights/

What Do Companies Do Today to Comply with Data Access Requests?

It is hard to develop hard and fast rules around what companies are doing and what they should be doing when looking across a broad range of industries. Certainly, companies with more data tend to have better systems. Most companies have dedicated teams to address privacy issues as well as privacy policies describing how they use data. There is not much evidence from our enquiries that suggest extensive disclosure of personal information to third parties. However very few corporates that we approached with requests, outside tech companies, appear to have systematized personal data in a way that privacy advocates would regard as best-in-class.

Future Directions – "Portability"

Traditional rules around data ownership tended to - at a minimum - give a person the right to access data held about them. Our anecdotal case study of one above shows that this is not a perfect system - sometimes it is difficult to access data, and certainly the data is not returned in any standardised format to allow easy reuse.

In response to this limitation, countries around the world have begun to draw up legislation around "portability" of data. For example:

- Article 20 of the EU's GDPR creates a "Right to Portability"[3]
- California's CCPA creates some limited rights of portability[4]
- Canada's Digital Charter contains portability within principle 4[5]
- India's PDPB contains a right to data portability[6]
- Australia's Consumer Data Right (CDR) is specifically targeted at portability[7]

The list above is certainly not exhaustive, and the specific rights do vary, despite similar language being used to describe the rights created. But there is certainly a trend in favour of so-called "portability" rights - essentially the right to obtain and direct electronic data about you, to another company of your choice. In theory, this will make reuse easier, opening up opportunities for data use across different settings to that originally intended when the data was collected or created.

In addition to this trend for general rights of portability, sector-specific portability regimes have been created, often where no general regime exists. This has been particularly common for the banking sector, often under the popular label of "Open Banking" (though again, this term has varied meaning). Successful sector-specific regimes may be a pathway to more general regimes - for example Singapore has been regarded as an open banking pioneer for some time[8] and is now consulting on reforms which include a broader data portability regime[9].

Australia's Consumer Data Right (CDR)

The CDR in Australia is a significant reform which is being rolled out sector by sector. Whilst often popularised as "open banking" - since banking was the first sector considered - the CDR goes far beyond banking. The next sectors listed for the rollout are energy and telecommunications. The clear intent, though, is that *every* sector of the economy will be included in due course - so all should be getting prepared for the CDR.

The CDR was sold on two main grounds: improving competition and enhancing opportunities for "digital innovation", notably local tech start-ups. We consider these below.

It would seem obvious that competition could be improved by streamlining application processes where they are seen as complex and difficult for consumers. It's expected that the CDR will make mortgage applications substantially simpler, for example - today these normally require time and effort from the consumer to extract statements or transactions from other accounts.

For other products the competition angle is harder to see. From a portability standpoint, it's already extremely easy to change an electricity provider, for example, requiring almost no data entry. Indeed, the only information required is usually some limited personal information for identification, and some basic information about the address, which is hardly taxing (and some form of ID will surely be required anyway for any CDR data transfer order). Telecommunication is similarly easy to sign up to, requiring little information.

What may improve competition is access to comparison services based on usage. All of these designated sectors today charge fees to consumers based on how the services are used. All of these sectors, therefore, generate rich usage data for a consumer, which is traditionally hard to access, and not naturally known by the consumer. By allowing usage data to be extracted easily in a standardised form, and by making product data similarly accessible, new forms of comparison or aggregation service will undoubtedly arise. This is perhaps the way competition will increase in the energy and telecommunication sectors.

Aside from aggregation and comparison sites, the "digital startup" justification has been touted. Certainly, it's reasonable to assume that more data will be a boon to the sector, which tends to sell itself on data-driven service offerings. It's also quite reasonable to assume some directly related ancillary services might be made possible via the collection of usage data. For example, your smart appliance might be able to switch itself on or off based on your patterns of electricity use and your billing plan, which could be fed from CDR data.

Other use cases that cross sectors are more challenging. This includes startups and traditional businesses. In particular, rich usage data is likely to be shown as predictive in a whole host of modelling or "AI" settings, but just because something "works" doesn't mean it should be done. Even within the currently designated sectors, this creates some interesting ethical questions for us to consider. Your electricity usage might well predict your chance of a loan default (the authors don't know either way), but even if it does, it doesn't necessarily follow that your electricity data should be allowed to be used to set your interest rate. We expect this sort of cross-sector data usage could cause significant social problems, as we outline below.

The current CDR rules require that this data be "reasonably needed" to provide the service in order to be requested – this is a core part of the "data minimisation principle"[10]. It's unclear exactly how broadly this term will be interpreted – certainly one could make an argument that many services could in theory exist without any data being provided at all. Our concerns below primarily relate to a broad interpretation of this term, which we suspect is the general intent of the regime given its stated goals.

Suppose, then, that we take a broad interpretation of the rules above, and allow your service provider X to ask you to "opt-in" to sharing some CDR data in order to get a "better deal". This could be your bank, electricity company, internet provider, or something else - it doesn't matter, the general argument is similar whatever the specific service or offer you consider. With this data, X decides whether you can receive their "better deal". If X's industry still requires the same total revenue pool (i.e. total costs and required profits are stable), then there is no way for X to offer such a deal in the long run, without a worse deal being given to the remainder, across the market as a whole. This is a basic form of market driven anti-selection which most actuaries should be familiar with.

The twist in this case is that CDR data is not universally held. There are people in society who do not have loans, or even bank accounts. There are people without good access to digital tools - access to which will likely be required to drive such offers. These people might be perfectly eligible for the "better offer" in theory (i.e. they might be low cost to service), but they are digitally excluded from it. Clearly, such people are also more likely to be vulnerable members of the community. It would seem unfair to ask such people to bear more cost than they currently do - this reinforces their current social exclusion. Many 'edge' cases exist - a women fleeing domestic violence may have little access to existing CDR household data, if they are in the husband's name. A migrant will not have existing product history. People might prefer to remain anonymous – and perhaps should have such a right without being penalised, in line with the general theme of APP2.

We suggest that the CDR (and similar regimes) will need to tackle this issue at some point - particularly if cross sector data sharing becomes commonplace. It must be recognised that data is not universally held, and that holders are generally already privileged members of the community. This creates a dramatic change from today's application forms for services, which ask people to enter information which can generally be known by anyone, or at worst easily found out by them. Arguments of competition improving consumer outcomes do not hold true if competitiveness is not constant, and digitally driven competition surely is not. Digital startups may give great and improved services to some members of society, but this probably should not come at the expense of those who already don't have internet access, and may already be struggling.

Our suggestion is that society must find a way to address this issue, without losing the benefits of the digital economy which regimes like the CDR can drive. We do not propose any solutions here for such a big challenge, but do suggest that it should not be ignored.

Conclusions and Challenges for the Next Decade

The next decade is likely to see continued change locally and across the world in the realm of privacy and data rights. Over the last few years, data portability has become increasingly incorporated into privacy and data regimes, and this is likely to become the norm over the next decade. With CDR as a very early example of this, Australia is likely to be ahead of international developments, and may be the first to seriously encounter some of the issues created by portability.

Of course, ownership and portability rights add things on top of existing privacy rights. As our case study of one showed, even the limited rights available today are not applied consistently. It is hard to imagine increased rights will improve matters, without some cultural change in the area, and greater public awareness. No doubt, if all consumers were demanding their data every day, Hugh would have received all of his data right away!

The challenge identified above of portable data not being universally available is not very well recognised but is likely to quickly become apparent if portability regimes are successful and data sharing is applied broadly. Disadvantaged people will suffer. We think this is an area to watch, particularly for consumer rights organisations.

A further interesting challenge will be one of discrimination. With richer datasets available for analysis, it is almost certain that some data will correlate with an attribute protected under anti-discrimination legislation, creating issues of indirect discrimination, even when all protected attributes are removed from the data. This has been well recognised in research circles for over a decade (see, e.g. Pedreschi et al. 2008[11]). Calls for reform in the area are likely to become louder as rich datasets become more and more available for analysis and prediction.

With all this change, we suggest that:

- actuaries, as users of data, should ensure they stay up to date on the changing landscape, through CPD or other activities; and
- actuaries, with our strong combination of technical grounding, business knowledge and commitment to the public interest, should have a strong voice in any debate on future standards and legislation.

We encourage all members to follow this area closely over the coming years, and contribute where relevant, either via the Actuaries Institute or other forums.

References and Notes

[1]https://www.oaic.gov.au/privacy/guidance-and-advice/what-is-personal-information/

[2]See https://www.abc.net.au/news/2020-01-23/australian-founder-of-clearview-facial-recognition-interview/11887112 or https://www.buzzfeednews.com/article/ryanmac/clearview-ai-fbi-ice-global-law-enforcement.

[3]https://gdpr-info.eu/art-20-gdpr/

[4]http://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article=

[5]https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00108.html

[6]https://www.prsindia.org/sites/default/files/bill_files/Personal%20Data%20Protection%20Bill%2C%202019.pdf

[7]https://www.accc.gov.au/focus-areas/consumer-data-right-cdr-0

[8]https://www.finextra.com/pressarticle/76370/singapore-tops-the-charts-in-open-banking-readiness-in-asia-pacific

[9]https://www.pdpc.gov.sg/guidelines-and-consultation/2020/05/public-consultation-on-personal-data-protection-(amendment)-bill

[10]https://www.accc.gov.au/system/files/CDR%20Rules%20-%20Final%20-%206%20February%202020.pdf paragraph 1.8 and related sections

[11]Dino Pedreshi, Salvatore Ruggieri, and Franco Turini. 2008. Discrimination-aware data mining. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '08). Association for Computing Machinery, New York, NY, USA, 560–568. DOI: https://doi.org/10.1145/1401890.1401959