Is data your most valuable asset that you never owned?

Peter Leonard[1]
Principal, Data Synergies

It is sometimes said that data is 'the new oil'. But appealing analogies can be misleading. There are many odd features of data that make it very different to oil.

Generally, data has little inherent value. Data value is derived not by *what data is*, but by *what can be done to create value* with data and then *endurably capture that value* by denying others the ability to do those things, while not also exciting regulatory intervention that may strip that value. Possible value depleting interventions include enforcement of competition (antitrust) laws, statutory creation of new 'consumer rights' over data, and enforcement by individuals of rights of access to, or portability of, transactional data (whether or not personal information about them) as held by data custodians.

Data can be infinitely reproduced and shared at effectively zero replication and sharing cost. Data does not derive its value through scarcity. Value in data may be created through investment in 'discoverability': in collecting and transforming raw data to enhance capability to link data to other data and then explore linked data sets for correlations. Often in data analytics projects about 75% of the cost is cleansing and transforming raw data to make it discoverable: the high-end work of then analysing the transformed data is the smaller part of the program budget. Discoverability often may be created within a privacy protected environment: in many cases, substantial data value can be created and commercialised whether or not particular individuals are identifiable.

Data may also 'fuel' value in other data: oil can't create value in other oil. And sometimes data derives value not through direct application of that data, but through enabling development of methodologies, algorithms and code for use on other data. AI didn't beat grand masters in chess and Go by being smarter, but by learning from 24x7x365 playing of games, generating 'training data' to inform machine learning. So data may enable code that then enables analysis of other data, in many cases then making that other data more valuable. And often a large volume of data of uneven quality can yield algorithms of substantial value which may then make poor data or narrow data sets more valuable. In short, data can be transformative in value of other data.

Valuation of so-called 'data rich' businesses is sometimes confused by failure to distinguish between the quantity and range of data sets that a business holds and the capabilities (or lack thereof) of a business to transform those data sets into actionable insights or other sustainable business advantage. Transformational methods and code and algorithms are often fungible across business sectors, with the result that data rich businesses concentrated within particular industry sectors may not achieve economies of scope that are available to cross-sector consultancies. Scarcity of human capital, and in particular experienced data

[1] Peter Leonard is a data, content and technology business consultant and lawyer and principal of Data Synergies. He was voted by peers as Sydney Technology Lawyer of the Year 2016 in *Best Lawyers Global 2016*. Peter chairs the Law Society of New South Wales' Privacy and Data Law Committee, the Australian Computer Society's Artificial Intelligence and Ethics Committee and the IoT Alliance's Data Access, Use and Privacy work stream. He serves on a number of relevant corporate and advisory boards, including the NSW Data Analytics Centre and the Australian Human Rights Commission's Human Rights and Technology Project. Peter was a founding partner of Gilbert + Tobin. Following his retirement as a partner in 2017 he continues to assist Gilbert + Tobin as a consultant.

scientists, means that much data that is captured today is not transformed and never achieves its potential value. Human capital remains the key investment in cleansing, transforming and linking data, in discovering useful correlations, and in creating and applying algorithms to data sets to derive actionable insights. Technology enables, but humans (still) create. And humans are ambitious, fickle and moveable - so innovative people culture will continue to be a key differentiator of good data businesses.

In short, the analogy of data as valuable 'oil' undervalues the value-adding contribution of the processes required to 'refine' data to power actionable insights for businesses. Good insights as outputs require great labour to create quality data inputs and to derive robust algorithms that are used as the engines of transformation. Which is one reason why many of the more ambitious predictions as to roll-out of applications of artificial intelligence have proven incorrect. And valuable business insights are often deployed in disrupted product or service sectors that are characterised by increasingly short product lifecycles, where returns on investment are highly uncertain. Markets for outputs of data are volatile and unpredictable. Actionable insights often have narrow application, a short shelf-life and require continuing innovation and reapplication. By contrast, refined (real) oil can be stockpiled. Oil is fungible across many industrial, transport and heating applications. The movement from fossil fuels to alternative energy is still agonisingly slow. Oil markets may appear to be volatile, but the market for outputs of data is no walk in the park.

Perhaps the most odd attribute of data is that (in most jurisdictions around the world) data cannot be 'owned', at least in the traditional legal concept of 'property'. Legally recognised 'property' may be tangible (chairs, dogs and pencils) or intangible (software, creative writing, trade marks and patents). Often a large component of intangible value is trade secrets, or as we usually call it in Australia, confidential information. Trade secrets are not 'property' in most national legal systems and in most (if not all) national variants of generally accepted accounting principles. Of course, the market capitalisation of both 'unicorns' and 'data giants' demonstrates that public financial markets and venture capitalists see value outside traditional classes of property. A single trade secret 'asset' can be worth millions, or billions, of dollars. Google emerged out of nowhere to dominate the search engine world using Google's trade secret algorithms. Google's success today depends upon protecting these trade secret assets collectively described as the Google brand. Many trade secrets derive their value through closely guarded central control: the recipe for Coke, the Google search ranking algorithm, and so on. These trade secret 'assets' may not appear in the balance sheet as assets, but derive value through being closely held, and by being closely held, scarcity is created and managed.

But increasingly data sets must be shared to some degree to yield value. Data sharing within multi-party data ecosystems is required to deliver almost all online services, and particularly internet of things (IoT) applications and many offline suppled products and services. Many IoT services, and online platforms such as Amazon and Alibaba, require a complex supply-side data sharing eco-system of five or more data holding entities to enable delivery of a service to an end-user and billing for that service. A business to consumer IoT service may include a retail service provider, a data analytics service provider, a cloud data platform, a telecommunications network services provider, a billing services provider, a mobile app provider and an IoT device provider, all sharing data without settled industry standards as to data minimisation and data security. In other words, at least some sharing of data is required to deliver many services, while at the same time the service provider should protect service value through imposition of safeguards and controls to ensure that 'the

service provider's data' (which it does not 'own' as 'property') remains defensibly trade secret and confidential.

So it should not be surprising that almost every SEC filing and information memorandum nowadays includes text such as:

> The loss of trade secret protection could make it easier for third parties to compete with our products by copying functionality. In addition, any changes in, or unexpected interpretations of, the trade secret and other intellectual property laws in any country in which we operate may compromise our ability to enforce our trade secret and intellectual property rights. Costly and time-consuming litigation could be necessary to enforce and determine the scope of our proprietary rights, and failure to obtain or maintain trade secret protection could adversely affect our competitive business position.

The fact that data is often about humans creates other complexities. Trust, ethics and social licence, and data privacy and data protection laws, also come into play.

Firstly, novel applications of data by businesses and governments are rapidly emerging at a time as reputable measures of consumer and citizen trust, such as The Edelman Harvard Trust Barometer, so that globally that citizen trust in business, government, NGOs and media are at historical lows. More pervasive uses and impacts of AI, 'unaccountable algorithms', and reported significant data beaches, can be expected to further erode consumer and citizen trust. And if businesses work from a proposition that consumers don't 'own', and therefore can't expect to control, transactional data generated about them, those businesses risk of further erosion of consumer trust. Perceptions of some data subjects of unfair distribution of benefits derived through data analytics may fuel more negative consumer reactions than concerns of other consumers as to their personal privacy. A perceived business imperative of sustaining consumer trust may become a more powerful constraint upon business behaviour than current privacy and data protection laws or emerging expectations of corporate social responsibility. Whatever the cause, in the last few years there has been an escalation in practical constraints as to acceptable uses of data about humans. This trend shows no sign of ending.

Second, data protection laws are rapidly expanding in jurisdictional coverage, reach and restrictions. Over 100 countries now have data protection laws. And human rights based systems of privacy laws – which are only some of the privacy regulating jurisdictions – are increasingly assertive in conferring rights of access, and sometimes 'control', of affected individuals to basic information about their activities and transactions. So even though an individual does not 'own' personal data about her or him, many national data protection laws confer significant rights upon data subjects to control how data about them is used and disclosed. And a right to take personal data elsewhere has some attributes of an ownership right. As an expression of an individual's 'human right' of 'personal autonomy', the European Union confers on individuals a right of portability of personal data about an individual's activities and transactions. Other jurisdictions, such as China, Russia, Indonesia and Vietnam, don't have similarly broad human rights in privacy, but are enacting intrusive requirements to keep personal data about their citizens in-country. These countries justify these 'data localisation' or 'data sovereignty' laws as necessary for national economic security, or a reasonable enabler of local businesses in competing with global businesses. Some jurisdictions, such as Australia and Hong Kong, confer on individuals rights of transparency as to collection and uses of personal information about them, but see an

alternative and more compelling reason to create a new consumer right in data. This is not a control right, or an ownership right, and it has no significant policy rationale in human rights. Rather, the right is envisaged as a lever for competition between service providers in industry sectors where complexity of data impedes product comparisons and switching between service providers.

In summary, data privacy laws now mix with competition, trade and national security laws, with markedly different weightings in different countries.

Data management decisions are also impacted by concerns as to trust, ethics and social licence, data security, protection of confidentiality of data and protection of data value.

It should not therefore be surprising that one of the hottest new disciplines is 'information governance': humans devising and delivering verifiably reliable technical, operational, security and contractual controls and safeguards around multi-party 'data ecosystems' in complex supply and distributions chains.

The developing Australian construct of a 'consumer data right' in relation to some basic transaction data sets in the banking sector – sometimes called 'open banking' – is unabashedly based in competition policy. The consumer data right will address information asymmetry between bank incumbents and competitors by empowering consumers to require on-provision of their transactional data sets. This right will facilitate analytics-driven 'apples with apples' comparison of features and cost of competitors' service offerings with features and cost of the service that a consumer is receiving from their incumbent provider. The right will also make switching easier when a consumer (maybe after reviewing that comparison) selects an alternative service provider.

Although the Australian construct of a consumer data right bears some similarities to the EU right of portability of personal data, the difference in policy basis for creation of the right will lead to fundamental differences in scope and implementation. But whether a consumer data right is conceived as a human right of personal autonomy, or as a tool for competition policy, there is an important commonality. The fact that data is not 'owned' by a data custodian in most jurisdictions enables robust regulatory initiatives. A legally enforceable right of a data custodian to deny others the use of (controlled but not owned) data may be variously created by contract, through application of principles of equity as to protection of confidential information, or by exercise of statutory rights in trade secrets. However, these are all legal rights to deny others *access to, or use of*, data: they are not rights *in* property.

Because in many (but not all) jurisdictions most data sets are not 'owned' by data custodians (as their legal property, such as through copyright as a form of intellectual property), it is easier for governments to change the bundle of rights and obligations associated with data. Altering rights attaching to legal property is cumbersome and expensive in many jurisdictions, as their national legal systems require the State to pay just compensation to adversely affected owners of property. Because data sets are not 'owned' by data custodians, the European Union can create a broad right for consumers of portability of their personal data from one service provider to another provider. The data custodian cannot assert legal rights of a property owner and the benefit of constitutional protection against unjust confiscation of property. The US tech giants loudly complain of European regulatory over-reach. However, their complaint is not as to violation of property rights, but as to unfair diminution of business value, and of EU regulatory long-reach (as the data may be collected and held outside Europe). And many EU data custodians are rightly concerned as to

inappropriate scope and scope creep if the portability right is allowed to extend beyond basic transactional data sets to include a right for consumers to demand (and give to a service provider's competitors) derivations such as imputed or analysed customer attributes computed by an incumbent service provider through investment in valuable intellectual property in algorithms and data analytics methodologies and processes.

So in this complex data-driven world, what is 'data value'? Data derives value to the extent that a data custodian has both the technical capability and legal right to:

- capture useful data points of sufficient granularity and number in a readily usable form, and
- bring together that data (i.e. make the data 'discoverable'), and
- transform and then analyse the data, to derive meaningful insights and to enable actionable decisions to be made by the data custodian (or by another person with whom the custodian shares the information); and
- through first mover or other advantages and operating under the cloak of confidentiality or trade secrecy safeguards buttressed by contractual provisions, deny others the ability to do those things, while not exciting regulatory interventions to confine their ability to capture value through exclusive control of that data and derivations from that data.

To summarise the current state of data science as applied to customer data:

- The value of data is derived through increased availability of data capable of aggregation and merger with other data sets to provide information in a form readily assimilated and used by humans (such as visualisations and other value-added presentations).
- Through data analytics methods of increasing sophistication, data acquires additional value. For example, algorithmically generated customer segmentation analyses and tools enable service providers to improve their ability to define and then target increasingly granular customer segments and to differentiate as to price and other terms offered to those customer segments.
- Data is also more readily available for analysis because it is more 'discoverable' as data taxonomies are standardised and as data extraction tools refined. There is more data that can be discovered and used. There is a rapidly expanding better range of already developed and tested data analytics tools and methods.
- Data is more readily available because businesses interact with each other and with consumers through increasing flows of consumer data and as each business is increasingly algorithmically driven in its own operations (therefore requiring better integration and availability of data across the business).
- Data value can be realised through improved ability of business to understand characteristics (attributes, preferences and interests, inter-relationships) of their customers and to infer characteristics of 'lookalike audiences' of prospective customers.

- Tools and methods for analysis and presentation of actionable insights are now widely readily available, including powerful tools readily available to consumers in the form of apps on smart phones or access to service comparison engines.

As new competitors and intermediaries emerge, every service provider will need to make, and regularly review, a number of the key strategic decisions about customer data that they collect.

- To what extent should the provider share the value that the provider can derive through data: for example, by empowering the provider's customers to make better, more informed decisions, and meeting demands by customers for enhanced access to data to facilitate customers analysing that data themselves?

- Should a provider respond only to regulatory compulsion, or should a provider strive to differentiate itself from its competitors by addressing demand by (some) customers for greater access to, and control over, data about them?

- Is it viable to keep customer data within a provider-controlled data ecosystem? To what extent will customer demand or regulatory intervention be such that certain customer data should be proactively made available into a more open data environment? If so, how much data, how made available, and to whom?

- How does a provider make data more available while adequately protecting its customers from security risks? In what circumstances should customers be taken to know and assume the consequences of such risks? More specifically, how does a provider verify that a customer properly understands the risks of conferring agency upon an intermediary to conveniently access sensitive financial data about a customer? How does a provider verify that an intermediary (such as a price comparison engine) purportedly authorised by a customer to access certain sensitive financial data about that customer is so authorised and will handle that data with due care?

- Where data is shared, and particularly where value is created through joint endeavour, is it jointly 'owned' (in the legally incorrect sense of control and ability to deny to others) and if so, what does this actually mean (given that business people, commercial lawyers within jurisdictions, and different legal jurisdictions, exhibit radically different views as to attributes of 'joint ownership')?

- What is ethical and fair, in a world where perceptions of fairness and social equity is increasingly contested, and social capital (and shareholder value) vulnerable to shifts in consumer sentiment and loss of citizen trust?

Achieving good information governance, and negotiating fair, balanced and durable data deals is not easy. This is why transparency of rights of use, and good information governance, will be a key differentiator of industry leading data businesses of the future, regardless of the trajectory and pace of development of global data privacy laws. And why competition and consumer protection regulators will be increasingly important stakeholders in data policy.

Service providers with corner-cutting or slapdash information governance processes will wither away through actions by private litigants or regulatory action, mistrust of business partners or erosion of consumer trust. The stakes are high: data driven businesses that wish to build sustainable data business models must do information governance really well.

Peter G Leonard
Principal, Data Synergies
Consultant, Gilbert + Tobin

M +61 411 089 003
E pleonard@datasynergies.com.au
LI https://www.linkedin.com/in/peleonard/