Big Data Fuels Digital Disruption and Innovation; But Who Owns Data?*

**Anthony Wong, Principal, AGW Lawyers & Consultants, anthonywong@agwconsult.com, and past Chair of the NSW Government ICT Advisory Panel*

1. INTRODUCTION

A staggering 2.5 quintillion bytes of data are created every day — with 90 per cent of the world's data produced in the last two years alone.[1]

Our economy has been moving from the physical world dominated by tangibles to one motivated by 'bits', 'waves' and 'droplets' of data (intangibles). Correspondingly, the values that define the wealth of our society are also shifting from the tangible to the Digital Economy. The First Industrial Revolution was powered by coal, but as we stand in the midst of the Fourth Industrial Revolution, innovation is now energised by the explosion of data from a myriad of sources.

A high percentage of 'Big Data' is generated by sensors, devices, smart meters and systems collecting transactional data, created largely through computerised and automated processes. Data also includes public and private sector data, and data concerning personal individuals.

With the world increasingly connected by the Internet of Things (IoT),[2] significant disruptions to long-standing business models and beliefs are taking place — from online shopping, connecting and communicating with Uber drivers, musing in driverless cars to transacting with digital currencies, all powered by data with the aid of applications and networks.

The use of intelligent software in conjunction with the rapid declining cost of digital storage are fuelling the assembly and combination of vast datasets of data for automated data processing and data mining. The algorithmic software, more cost effective and efficient than human readers, is being progressively deployed across all domains of our society. Data mining will unlock and discover new forms of value, connect previously unseen linkages and provide insights to stimulate growth and

[1] IBM, *Bringing Big Data to the Enterprise* <http://www-01.ibm.com/software/au/data/bigdata/>.

[2] A phrase coined by Kevin Ashton to describe the system where devices are connected through the Internet to transmit, compile and analyse data.

innovation in the Digital Economy. Increasingly, copyright and non-copyrightable materials are being used as 'data feed' to 'fuel' text and data analytics.

As our society's dependence on the Digital Economy increases with the rapid evolution of 'Big Data', it has heightened the issues of 'propertisation' and 'commoditisation' of data. Although the debate on property rights in data is not new, the issue has taken on a renewed emphasis in the context of 'Big Data'. This debate has centred on the ability and freedom to use and extract value from data in the endeavour to ascertain insights to new discoveries, innovation and economic growth.

Protecting value and proprietary rights in 'Big Data' involves a balancing act between the many vested interests, including the interests of the 'purported' owner, the 'custodian', the interests of competing third parties, and the interests of the public to access and use data. Many see 'Big Data' as a new commodity — a form of currency — just like spices were in the days of the spice trade in the East.[3]

As identified in the White House report, *Big Data: Seizing Opportunities, Preserving Values* ('White House Big Data Review'),[4] data is viewed as a major source of value and economic activity. The report concluded that the explosion of data in today's world can be an unprecedented driver of social progress but the challenge lies in understanding the many different contexts in which data comes into play including data as property (who owns it), data as a public resource (who manages it and on what principles), and data as identity or as an expression of individual identity.[5]

Economies are starting to form around data, irrespective of whether an adequate legal framework has been built around it. For the most part, traditional intellectual property laws and related rights have proven to be an inadequate solution. These limited interests in data leave out several important rights, which are pertinent in the world of

---

[3] Frank J Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons, Inc, 2012); John Lucker, 'Big Data Alchemy: Turn Info Into Money Data Markets', *Information Week* (online), 13 May 2013 <http://www.informationweek.com/big-data/big-data-analytics/big-data-alchemy-turn-info-into-money/d/d-id/1109933>.

[4] Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (May 2014), 9, <http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf> ('White House Big Data Review').

[5] Ibid 3. The report referred to a quote by Professor Sheila Jasanoff, Harvard Professor of Science & Technology Studies.

'Big Data', including the right to control access, disclosure and use. The debate on data ownership rights has intensified as the use and control of data assets become more and more critical to our future economies and our ability to innovate — requiring re-balancing of the commercial, private and public interests in data, and not least, privacy concerns.

1.1 What are the sources and types of 'Big Data'?

An understanding of the sources and types of 'Big Data' is required to properly comprehend the debate on ownership rights in data. Aggregators of 'Big Data' include social media, the web, geospatial data, web-enabled and wearable devices, Internet of Things (IoT), sensors, systems capturing transactional data and audit logs.

The convergence of technology and telecommunication have facilitated the proliferation of Internet-enabled device sensors and devices which in turn have provided the base-capacities to collect data from millions of individuals. Personal location data can also come from GPS devices, cell-tower triangulation of mobile devices, mapping of wireless networks and in-person payments.

Data comes in various forms and includes public and private sector information, and data about personal individuals. Public sector information (PSI) comprises the largest subset of 'Big Data' and may consist of research data, public registers, train and transport timetables and administrative data from federal, state and local authorities.[6]

PSI is created based on specific laws and regulations and includes company registers, intellectual property registers, land and titles information, land survey and geospatial data, vehicle registers, population and census data. Such data has traditionally been used by institutions of government for public and government administration and policy-making. With the advent of 'Big Data' and open access facilitated by internet and cloud technologies, it may now be readily available to the private sector. However, release of PSI may be restricted due to privacy and security concerns and other overriding public interest against disclosure. Access to PSI is typically regulated by

[6] For further information, refer to the European Commission Directorate General for the Information Society, *Commercial Exploitation of Europe's Public Sector Information – Executive Summary* (2000) <ftp://ftp.cordis.europa.eu/pub/econtent/docs/2000_1558_en.pdf>.

statutes but may also be accessible on the basis of Freedom of Information laws.[7]

Private sector information includes datasets containing customer lists, transactional information covering a spectrum of financial, payment, purchasing (both offline and online) and service transactions.

As described in the report by the US President's Council of Advisors of Science & Technology, some data is 'born digital, created specifically for digital use by a device or data processing system.'[8]

History will see the 'Big Data' revolution as disruptive and another big game-changer since the invention of the printing press and the internet, challenging centuries-old business models. In the world of 'Big Data', these datasets could be created, collected and obtained (sometimes even verified) automatically or as a secondary by-product of another business enterprise. Some will require the investments of time, capital and labour, while others may only require computing processing time. It all comes down to the types and forms of datasets, how they are derived and the purpose they serve.

The challenge here is — should some of these datasets be publicly available, or should some producers of capital and labour-intensive datasets be able to inhibit and restrict access to these datasets from data mining and analysis and to safeguard their ability to recoup their investments?

1.2 The structure of the chapter

This chapter explores ownership interests in data assets in the era of 'Big Data' and goes beyond intellectual property law. As discussed by Nimmer,[9] it includes rights developed in areas of law not commonly viewed as property in our law regimes.

In writing this chapter I have been taken on a journey, and sharing Merges' sentiments in the opening pages of his book,[10] it is perhaps a longer journey than anticipated. Evolving property law to the next level to parallel the challenges posed by 'Big Data' –

---

[7] See Thomson Reuters, *New South Wales Administrative Law* (at 25 April 2016) [50.120]; see also Thomson Reuters, *Federal Administrative Law* (at 25 April 2016) [FOI.0.10].
[8] White House Big Data Review, above n 4.
[9] Raymond T Nimmer, Thomson Reuters, *Information Law* (at 7 June 2014) [2:1].
[10] Robert P Merges, *Justifying Intellectual Property* (Harvard University Press, 2011) 4.

taking the evolutionary steps from real property to embrace intellectual property, and then to 'Big Data', is perhaps over-stretching the traditional concepts of property law; it definitely has 'the feel of a northern fir in the tropics, or a damp fern in the high desert.'[11]

However, I take comfort in the words of Nimmer, 'We deal here with a major, transformative phenomenon. We do not need to capture the Genie into a single, confined bottle in order to discuss its consequences in our lives and in the life of the law.' [12]

This chapter takes up one of the challenges identified in the White House report — 'Big Data' as property (who owns it); and the layered complexities and issues pertaining to the granting of property rights in data. In its relentless 'technological progress', the 'Big Data' phenomenon has overtaken the slow march of our law and has embraced and encapsulated some of the facets of our concepts of property without giving due regard and serious thought to the implications of treating data as property. In an attempt to create order from a run-away phenomenon, this chapter will review whether there should be underlying policy reasons to accord some form of 'property rights' in the context of 'Big Data', and if not some 'bundles of rights'.

In Section 2, we review the state of play of property rights in data and the legal and economic structures that define ownership of data.

In Section 3, we explore the challenges posed by data ownership and in Section 4, we briefly outline the concept of control versus ownership.

This chapter does not purport to resolve all of the issues that relate to 'Big Data' and 'property rights', but provide a starting base for understanding some of the emerging challenges of ownership and control of data. It will touch on but not cover in any detail the traditional bases for intellectual property protection of information such as copyright, patent and the quasi-intellectual property doctrines, under confidential information and trade secrets. These areas have already been exhaustively covered

[11] Ibid.
[12] Nimmer, above n 9, [1:8].

elsewhere.

2. STATE OF PLAY ON PROPERTY RIGHTS IN DATA (INCLUDING ECONOMICS OF DATA) IN THE CONTEXT OF 'BIG DATA'

In a study of the economic value of open data, the McKinsey Global Institute determined that government data could unlock more than US$3 trillion in value every year in seven domains of the global economy: education, transportation, consumer products, electricity, oil and gas, health care and consumer finance.[13]

The EU Commission has also launched an Open Data Strategy for Europe, which is expected to deliver a €40 billion boost to the EU's economy each year.[14]

To fully appreciate the economics of 'Big Data', we should begin by exploring the benefits of 'Big Data'.

2.1 What are the benefits of 'Big Data'

Data's value lies in its use, not its mere possession.[15] 'Big Data' tools allow us to combine, interrogate, mine and analyse large structured or unstructured, multiple datasets[16] with ease where the sum of these datasets is more valuable than its parts; allowing us to triangulate and identify correlations that were not easily done previously. It is a competitive advantage to find new ways to interpret data and process them faster using 'Big Data' analytics. '[J]ust as with gold, it is through mining that one finds the nuggets of value that can affect the bottom line.'[17]

'Big Data', including data about individuals, also has value in the marketplace — our

[13] McKinsey Global institute, *Open Data: Unlocking Innovation and Performance with Liquid Information* (2013) <http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information>.

[14] European Commission, '*Digital Agenda*: Turning Government Data into Gold' (Press Release, 12 December 2011) <http://ec.europa.eu/digital-agenda/en/news/turning-government-data-gold>.

[15] Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Eamon Dolan/Houghton Mifflin Harcourt), 2013) 122.

[16] In addition, data in its raw form is now also available for analysis, in contrast with traditional structured databases.

[17] Ohlhorst, above n 3, 19.

age, sex, our genes, where we live, our ethnicity, education, what we do for a living, our financial and transaction records, our beliefs, preferences, purchasing and lifestyle habits — all provide correlation data as to the products and services that we might be interested in, providing valuable data for targeted dollars and marketing.

The roles of institutions are also changing as we clamour for ever-greater productivity and efficiency. In their endeavours to comply with money laundering legislation, including the 'Knowing your customer' requirements, financial institutions have discovered that they have been sitting on a gold mine of data about their customers including:

- Where they shop
- What they buy and their favourite restaurants
- Where they go for holidays
- Facilitating their move to provide new 'non-financial' services including travel and car insurances and by partnering with traditional and non-traditional players.

Similarly, telecommunications providers have been collecting data as part of their 'normal' business enterprise including:

- Our service payment transactions and our service usage plans
- Our movements and places where we visit — using the geo-location tracking abilities in our mobile devices
- Facilitating their ability to promote targeted marketing products and services based on our geo-location
- And our subscriber content and viewing habits.

Much of the value of 'Big Data' may be derived from the secondary uses of data, rather than its 'normal' or primary use.

'Big Data' also challenges existing privacy frameworks and principles including the definition of personal information, principles of data minimisation, purpose limitation and the concept of consent.

Many recent commentators are of the view that anonymisation or de-identification is not robust enough against future re-identification techniques that are being developed for many legitimate applications of 'Big Data'.[18] 'Big Data' allows us to data mine, analyse and create profiles on individuals easier than at any time in the past.

Although many scholars from the American, English and European traditions have advocated for the recognition of property rights in personal information, the law is reluctant to accept the subsistence of fully-fledged proprietary interests in data as such.[19]

This chapter does not cover property rights in personal information in any detail. The subject of property rights in personal information is complex and deserves a chapter in its own right. Numerous authors have covered this subject exhaustively, including Patricia Mell in her paper, 'Seeking Shade in a Land of Perpetual Sunlight: Privacy as Property in the Electronic Wilderness'[20] and also Vera Bergelson, 'It's Personal but Is It Mine? Toward Property Rights in Personal Information'.[21]

Hemnes, in his analysis on the subject,[22] questions whether 'identity theft' laws have endowed some form of proprietary interests in personal information. 'If something is capable of being stolen, and if one can remedy the theft, does this not imply that it was owned in the first place?'[23] He concluded that there are certainly forces pushing in that direction, but it has not happened yet and suggested that in a future period,

---

[18] President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective* (1 May 2014), 38, <http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf>.

[19] Maurizio Borghi and Stavroula Karapapa**,** *Copyright and Mass Digitization: A Cross-Jurisdictional Perspective* (Oxford University Press, 2013) 147.

[20] Patricia Mell, 'Seeking Shade in a Land of Perpetual Sunlight: Privacy as Property in the Electronic Wilderness' (1996) 11(1) *Berkeley Technology Law Journal*.

[21] Vera Bergelson, 'It's Personal but Is It Mine? Toward Property Rights in Personal Information' (2003) 37(2) *UC Davis Law Review* 379 at 412.

[22] Thomas Hemnes, 'The Ownership and Exploitation of Personal Identity in the New Media Age' (2012) 12(1) *John Marshall Review of Intellectual Property Law* 29.

[23] Ibid.

personal information in one's digital identity 'will have grown into something closer to a property right in one's identity.'[24]

Most jurisdictions in Australia have enacted legislation in relation to identity theft.[25]

The statutory provisions may have countered the threat of 'identity theft' but it has not enlightened us or justified the basis of the protection using any property law concepts behind the 'thing' – the identity, that is capable of being stolen.

2.2 Notions of property and data

The foundation of our Australian legal system is property law and even in the 21st Century, ownership of and property rights are still of paramount concern. The law of property has developed over centuries, originally for the purpose of protecting interests in land and, relatively recently, interests in other things where rights of ownership may be exerted such as goods, and intangible properties such as rights under intellectual property.

Property rights evolve and change to address certain practical needs of a given epoch in our society. Those needs change alongside our changing values and norms. Literature abounds on the different senses in which the term 'property' has been used to encapsulate the move from the traditional notions of property, such as land and chattels, to the notion of property in intangibles. We are embarking on yet another significant leap, this time specifically, property or 'property-like' considerations in 'Big Data'.

It is difficult to define property with any precision as the 'notions of property inevitably change to reflect their context.'[26] Property deals with rights and if recognised under

---

[24] Ibid.

[25] See, eg, *Law and Justice Legislation Amendment (Identity Crimes and Other Measures) Act 2011* (Cth); *Criminal Code Act 1995* (Cth); *Crimes Amendment (Fraud, Identity and Forgery Offences) Act 2009* (NSW).

[26] Huw Beverley-Smith, *The Commercial Appropriation of Personality* (Cambridge University Press, 2002) 286.

established heads of law are claims 'good against the world', often described as 'rights to exclude others'.[27]

Data is another dimension to the intangible properties such as rights under intellectual property. Data shares a number of characteristics with intellectual property including the following:

- It may be contained in a tangible medium like paper, document, DVD, computer, device or tape.
- It cannot be physically possessed.
- The same data may be included in many different products.
- It may be the copyrightable work itself as contained in an expression.
- It may be the confidential information subject to secrecy or confidential arrangements.
- Data may be contained in a patent.
- The provision, use or transfer of data does not exhaust it in the true real property sense.

2.3 Protection by existing intellectual property and related rights

Existing laws in relation to copyright, patent, confidential information and trade secret, and trademark, all relate to and protect rights involving information.

As observed by Nimmer, '[C]opyright law has become a primary source of property rights in information in the 1990s.'[28] However, copyright is not an adequate framework for the consideration of property rights in all droplets of data as it only provides

[27] See Merges, above 10, 100.
[28] Nimmer, above n 9, [2:8].

owners with a limited property right in the expression of the information.[29] Copyright law does not concern itself with the control or flow of ideas, facts or data per se. The data components contained in the copyrighted work may not be protected, no matter how valuable. Ideas and facts are generally regarded to be in the public domain.[30]

The *Copyright Act 1968* has been expanded and extended to protect technological protection devices that control access to copyright material.[31] However, they leave out important rights, which are pertinent in the world of 'Big Data', including the right to control use, except to the extent provided by the exclusive rights under the *Copyright Act*.[32]

There is currently much debate around the very concept of 'use' as illustrated by the long-running Google Books litigation where US federal Judge Chin ruled that the project is within the bounds of US copyright law.[33] Google had not secured permission from copyright owners to digitise their books and/or to display short 'snippets' of surrounding text in relation to the search term online.

Borghi, in his analysis of the concept of 'use' (in copyright), suggests that it 'involves various kinds of activities, some of which are free for anyone to carry out — reading, listening to, or viewing a work and enjoying it, learning from it, and building upon it — and some others are reserved for the author.'[34] In the context of Google Books, he added that the:

> Technological transformative uses, by contrast, include activities where the work is no longer used as a work but as something else — for instance, as a

[29] The nature of the copyright in a literary, dramatic or musical work is defined under the *Copyright Act 1968* (Cth) s 31.
[30] See Pamela Samuelson, 'Is Information Property?' (1991) 34(3) *Communications of the ACM* 16.
[31] *Copyright Act 1968* (Cth) ss 116AN, 116AO and 116AP.
[32] Australian Law Reform Commission, *Copyright and the Digital Economy – Final Report* (November 2013) <http://www.alrc.gov.au/publications/copyright-report-122>. See Chapter 11 ('Incidental or Technical Use and Data and Text Mining'), which deals with and advocates for fair use or dealing for text and data mining for non-commercial research and non-consumptive use.
[33] *Authors Guild Inc v Google Inc* , 954 F.Supp.2d 282 (S.D.N.Y. 2013).
[34] Maurizio Borghi and Stavroula Karapapa**,** *Copyright and Mass Digitization: A Cross-Jurisdictional Perspective* (Oxford University Press, 2013) 45.

> carrier of data to feed information into computers. This is what we call, using an umbrella term, automated text processing.[35]

It is unlikely that the 'technological transformative uses' and the large-scale digitisation by Google in the US would be supported under current Australian copyright law. As suggested by Nimmer, 'It is also true that copyright does not generally create any right to prevent others from using or disclosing the facts or expression to others if this occurs without copying the expression.[36]

The right to control use of information may also arise under patent or other laws. Patent protects the use of ideas or information contained in the patent, by restricting the practice of the invention for a period of time.

In Australia and elsewhere, whether information can be properly characterised as property in the context of confidential information, have been subjected to much academic and judicial commentary over the last half century.[37]

However, if the owner of the confidential information places it in the public domain and accessible for 'Big Data' mining and analysis, the inherent 'secrecy' may be lost.

Confidential information is sometimes described as having a proprietary character, not because property is the rational basis of protection, but because of the effect of equitable protection.[38] In Australia, as in the United Kingdom, there is authority which supports the proposition that information is not property.[39]

---

[35] Ibid.

[36] See Raymond T Nimmer and Patricia Ann Krauthaus, 'Information as a Commodity: New Imperatives of Commercial Law' (1992) 55 *Law and Contemporary Problems* 115.

[37] For an introduction to the protection of Information using the law of confidential information, see Lahore, LexisNexis, *Patents, Trade Marks & Related Rights* (at 25 April 2016) [30,000].

[38] Lahore, LexisNexis, *Patents, Trade Marks & Related Rights* (at 25 April 2016) [30,020].

[39] See, eg, *Federal Commissioner of Taxation v United Aircraft Corp* (1943) 68 CLR 525 at 534; *Moorgate Tobacco Co Ltd v Philip Morris Ltd (No 2)* (1984) 156 CLR 414 at 438; *Breen v Williams* (1996) 186 CLR 71 at 81, 90, 111, 125; and *Australian Broadcasting Corporation v Lenah Game Meats Pty Ltd* (2001) 208 CLR 199 at 271.

3. CHALLENGES TO THE CURRENT CONCEPTS OF DATA OWNERSHIP

The rapid emergence of 'Big Data' and our society's dependence on the Digital Economy have heightened the debate on our ability/freedom to use and extract value from data, compilations and datasets without fear of prosecution in the endeavour to provide insights to new discoveries, innovation and growth.

Granting separate property rights to discrete and droplets of data (datasets) would create a substantial barrier to the evolution of 'Big Data' and our ability to mine valuable gems from these datasets.

In the world of 'Big Data', these datasets could be created, collected and obtained (sometimes even verified) automatically or as a by-product of another business enterprise. Some will require the investments of time, capital and labour, while others may only require computing processing time. It all comes down to the types and forms of datasets, how they are derived and the purpose they serve.

The different types and forms of 'Big Data' have and will continue to challenge our thinking and concepts around the question of data ownership. It has also created uncertainty in the boundaries of control and data ownership.

Rights in data come in many forms and from a variety of sources. For the most part, traditional intellectual property law has proven to be inadequate to provide protection.[40] The traditional intellectual property regimes do not provide adequate cover for data and information-based products sufficiently well. Indeed, these laws would exclude most of the 'Big Data' datasets (in whole or in part) from protection.

A large percentage of 'Big Data' datasets (including outputs from device sensors, mobile and GPS devices, smart meters, systems collecting transactional data) created through a largely computerised and automated process will suffer serious hurdles in securing copyright protection and will probably not have sufficient 'authorial' contribution for copyright to subsist.

---

[40] See Kristen Osenga, 'Information May Want to Be Free, But Information Products Do Not: Protecting and Facilitating Transactions in Information Products' (2009) 30(5) *Cardozo Law Review* 2099 at 2101.

In addition, due to the dynamic nature of some 'Big Data' databases, a number of contributors and researchers, analysts and programmers could also be involved. As in *IceTV Pty Ltd v Nine Network Australia Pty Ltd*,[41] and *Telstra Corp Ltd v Phone Directories Co Pty Ltd,*[42] there could be difficulties in identifying authors where a work is created and updated over time by a large number of people and using automated processes which lack authorial contribution.

In *Phone Directories*, Gordon J found that none of the 'authors' exercised 'independent intellectual effort' or 'sufficient effort of a literary nature' in creating the phone directories[43] and emphasised that the authorship/originality nexus is important in subsistence of copyright.[44]

The current copyright regime does not protect facts, compilations/databases generated by computers with little or no authorial input or insufficient 'originality' as illustrated in *IceTV* and *Phone Directories*.

Many datasets or databases that are by-products (ie. secondary) of an organisation's main activities (such as airlines schedules, stock market data, member directories, box scores, real estate listings and results of scientific experiments) would probably not receive protection, unless there is sufficient 'originality' to meet the threshold enunciated in *IceTV* and *Phone Directories*.

Although sufficiency of 'originality' for copyright subsistence in a form of expression is a matter of fact and degree,[45] the author/originality correlation poses great challenges for those seeking to protect 'Big Data' datasets which are created or collected with little or no authorial involvement. In fact, most 'Big Data' aspirants strive for data comprehensive over selectivity.

Using vast computing resources and power including cloud computing, algorithmic

---

[41] *IceTV Pty Ltd v Nine Network Australia Pty Ltd* (2009) 239 CLR 458.
[42] *Telstra Corp Ltd v Phone Directories Co Pty Ltd* (2010) 194 FCR 142. The High Court refused to grant Telstra special leave to appeal the decision of the Full Federal Court.
[43] *Telstra Corporation Ltd v Phone Directories Pty Ltd* [2010] FCA 44, [340].
[44] Ibid [344].
[45] *IceTV Pty Ltd v Nine Network Australia Pty Ltd* [2009] HCA 14 at [99]; *Telstra Corporation Ltd v Phone Directories Pty Ltd* [2010] FCA 44 at [25] and [26] per Gordon J.

programs are more cost effective and efficient (both in time and money) during data mining and analysis in identifying trends and correlation in 'Big Data'. These algorithmic programs may have been imbued with the requisite selective logic by their programmers, which effectively replace the use of human authors in the selection, identification, verification and presentation of the expressed output.

In *Phone Directories*, Gordon J found that there was no relevant 'intellectual effort' by Telstra's employees or contractors in understanding or applying the rules which determines the form of expression of the phone directories (Rule). The relevant 'intellectual effort' was expended on the computer systems and the development[46] of the Rules[47] used to generate the phone directories and was anterior to the phone directories taking its material form. Her Honour was not prepared to carry through the 'intellectual effort' expended on the computer systems and the development of the Rules to the phone directories. The authorship/originality nexus was broken.

Without legislative reform, our 19th century copyright concepts of authorship and fixation, will pose challenges for those seeking to protect their datasets in the era of 'Big Data'. Unless the law is amended, investments in a large percentage of 'Big Data' datasets will be left without effective protection. However, the lack of protection for 'Big Data' datasets would be beneficial for those advocating for narrower ownership rights in data.

3.1 Issues pertaining to Computer Generated databases and datasets

Commentators have long distinguished between computer-assisted[48] and computer-generated works. Under current Australian law, the former category posed few copyright problems, but computer-generated 'Big Data' datasets with little or no

[46] *Telstra Corporation Ltd v Phone Directories Pty Ltd* [2010] FCA 44, [165].
[47] The Rules determine the form of expression of the phone directories and where human discretion was exercised, it was dictated by the Rules. See *Telstra Corporation Ltd v Phone Directories Pty Ltd* [2010] FCA 44, [162]–[166].
[48] Here the computer is used as a tool equivalent to the painter's brush or the writer's pen by the author in the creation of the work.

human intervention would be a stumbling block to copyright's subsistence.

Using the reasoning of Gordon J in *Phone Directories,* an algorithm written for 'Big Data' mining and analysis by a separate person from the person executing the algorithm would break the authorship/originality nexus, resulting in no protection for the selected and resultant dataset from the data mining and analysis.[49]

Our courts have taken a rather literal and technical approach in *IceTV* and *Phone Directories* in relation to the roles and uses of computers and their equivalency to the painter's brush or the writer's pen in the creation of the databases by the author. It may be unfortunate for some that the courts' focus was on the 'last step in the creative process and then look for an immediate, direct link to a computer or human in order to determine authorship'[50] without taking a wider view of the whole creative process.

The UK has implemented specific provisions to protect computer-generated work.[51] Section 9(3) of the UK *Copyright Designs and Patents Act 1988* provides that 'in the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work were undertaken'.

Despite the recommendations of the Copyright Law Review Committee[52] and the Copyright Law Review Committee[53] to adopt the above UK's approach, no legislative changes have been undertaken by the Australian Government, resulting in the

---

[49] See the statement by Gordon J: 'the person or persons who utilise the Rules and who, therefore, are submitted by the Applicants to be authors of the Works, do not exercise either 'independent intellectual effort' or 'sufficient effort of a literary nature' to be considered an author within the meaning of the Copyright Act': *Telstra Corporation Ltd v Phone Directories Pty Ltd* [2010] FCA 44, 162.

[50] Lief Gamertsfelder, *Corporate Information and the Law* (LexisNexis Australia, 2013) 40 [2.47].

[51] Section 178 of the UK *Copyright Designs and Patents Act 1988* defines 'computer-generated work' to mean work 'generated by computer in circumstances such that there is no human author of the work'. Similar provisions have been replicated in New Zealand, Ireland, India, Hong Kong and South Africa. The US, while not having enacted similar provisions, has been able to better protect computer-generated work.

[52] The Copyright Law Review Committee (CLRC) in 1995 released its report on copyright and computer software. The report included a section on computer-generated works.

[53] See paragraphs 5.45 and 5.46 of the report by the Copyright Law Review Committee, *Simplification of the Copyright Act 1968, Part 2 — Categorisation of Subject Matter and Exclusive Rights, and other Issues,* Commonwealth of Australia (1999).

situation whereby the wider use of technology will hinder the securing of copyright protection in compilations and datasets including computer-generated works. Perhaps that is a blessing in disguise for the proponents of the information commons.

There is no *sui generis* database right in Australia, such as in the EU Database Directive in the EU.

In the era of 'Big Data', rights to control access, disclosure, use and the ability to extract value from data are paramount. With the advances in technology and the rapid proliferation of cloud and internet technologies, wholesale reproduction or copying as a step in the process of data mining and analysis may be a thing of the past. 'Big Data' tools are now coming to maturity that will allow the combination, interrogation, mining and analysis of large structured or unstructured, multiple datasets with ease without the need to undertake wholesale reproduction or copying as a requisite step.

4. THE CONCEPT OF CONTROL VERSUS OWNERSHIP

In the era of 'Big Data', the right to control access to the data contained in a system can often be more important than the right to control copying of the data.

Nimmer, in his paper, expanded the concept of control — '[t]he right to control another's access to information can implicate several distinct bodies of law, including the law of trade secrets, criminal law, communications law, and various laws relating to privacy interests.'[54] Copyright also plays a role. In the context of privacy law, the concept of control commonly referred to as 'notice and consent' is illusory in the world of 'Big Data'.

I have provided two examples below to illustrate the complexities inherent to the question of ownership and control of information.

[54] See Nimmer and Krauthaus, above n 36, 118.

4.1 Example 1 – Unauthorised Access to Data

Computer crime laws treat unauthorised access to computer systems in a manner analogous to trespass. The concepts commonly associated with the law of torts and property — the action for trespass to land (real property) and to chattels have been extended to protect data in the computer environment by a number of state and federal statutory criminal provisions in Australia. Rights to control access to information are found in the Australian federal cybercrime provisions— in the *Criminal Code 1995* (Criminal Code) and in the equivalent state provisions.[55]

The *Criminal Code* establishes a computer system as a protected environment and creates a crime analogous to trespass to a person's home – for unauthorised access. Even if no direct harm occurs, s 478.1(1)[56] recognises that the provider of the protected computer system has an inherent right to exclude unauthorised access to data held and restricted by an access control system associated with a function of the computer.[57]

The access control right in s 478.1(1) of the *Criminal Code* requires the presence of an 'access control system associated with a function of the computer'— i.e. a protected zone or environment. As in trespass to land, the section also requires that the act of unauthorised access to the data held in a computer be intentional. However, there is no requirement that the data be confidential in character or be owned by the provider of the protected computer system.

The above is an example of government intervention using the statutory framework to protect proprietary interest (the data) in a digital environment, without the requirement to establish ownership in the data held and restricted by an access control system associated with a function of the computer.

The right accorded by the *Criminal Code* is a claim 'good against the world', providing a 'right to exclude others', and is strikingly similar to what the provider of the protected computer system would have enjoyed if they own the data in the protected computer

---

[55] Australia adopted the *Council of Europe Convention on Cybercrime* on 1 March 2013.
[56] *Criminal Code Act 1995* (Cth) s 478.
[57] Ibid s 478.1(3).

system.

In addition, the *Copyright Act* may also be invoked if the 'access control system'— technological protection devices that control access to the copyright material — has been circumvented.[58]

4.2 Example 2 — Personally Controlled Electronic Health Records

One of the areas that will greatly benefit from 'Big Data' is health care. Using 'Big Data' tools to combine, interrogate, mine and analyse large structured or unstructured and multiple datasets, one would have the ability to triangulate and identify correlations in health data to discover a cure for a disease, sickness or to improve medical knowledge.

The Australian Government's personally controlled electronic health (eHealth) record system[59] was launched on 1 July 2012 utilising the framework provided by the *Personally Controlled Electronic Health Records Act 2012* ('PCEHR Act').

People seeking healthcare in Australia ('Patient') can register for an eHealth record — a secure, electronic summary of their health information such as prescribed medications, allergies and treatments that they have received. The eHealth record system gives the Patient control over their health information and they can choose to 'opt-out' at any time. The Patient controls who has access to their eHealth record, what information others can see and what records are uploaded by establishing access controls on their eHealth record.

Certain documents may be marked as 'Restricted Access' by the Patient and the Patient may provide the 'Restricted Access' code to the healthcare providers that they want to have access to the 'Restricted Access' documents in their eHealth Record. The Patient may also control the level of read and write access to their eHealth record by each healthcare provider.

---

[58] *Copyright Act 1968* (Cth) ss 116AN, 116AO and 116AP.

[59] See Department of Health, Australian Government, *Welcome to My Health Record* <http://www.ehealth.gov.au/internet/ehealth/publishing.nsf/content/home>; and Office of the Australian Information Commissioner, *My Health Records* <http://www.oaic.gov.au/privacy/privacy-act/e-health-records>.

The above implementation of the personally controlled electronic health record system poses many legal questions.

At the CeBIT's eHealth conference in Sydney in May 2014, Professor Morgan[60] raised concern about the issue of ownership and control of the personally controlled electronic health record system.

Are we able to mine the personally controlled electronic health record system for the 'nuggets of gold' that could uncover the cause of a particular disease, a cure for a sickness or a particular lifestyle that might trigger a certain illness?

Doctors in Australia have fiercely guarded their ownership and control over their patient records. The High Court in the case of *Breen v Williams*[61] established that patients have no proprietary rights or interests in the information contained in their own medical records.

The conclusion was based on the purpose for which the medical record was created, as well as the nature of the particular relationship between the patient and the health care professional. The High Court concluded that a medical record is ordinarily created by the doctors or healthcare providers for their own professional purposes in order to provide health care services to the patient, including diagnosis, treatment and advice. As the creation of medical records is not generally a term of any contractual relationship between a patient and the doctor or healthcare provider, medical records, as a general rule, are the property of the doctors or the healthcare providers that created them.

However, the law has not granted doctors and healthcare providers exclusive rights to patient information. Doctors and healthcare providers owe their patients a legal and

---

[60] Professor Kenneth Morgan, special advisor to the Vice-Chancellor of Flinders University on cyber-security and resilience, discussed the future of e-health at CeBIT's eHealth conference in Sydney in May 2014. As reported by Cynthia Karena, 'Ownership of Patient Records, Just One Challenge in E-Health', *Sydney Morning Herald* (online), 7 May 2014 <http://www.smh.com.au/it-pro/government-it/ownership-of-patient-records-just-one-challenge-in-ehealth-20140507-zr66p.html>.

[61] *Breen v Williams* (1996) 186 CLR 71; Dawson and Toohey JJ at 88–90 expressed the view, in the context of the information in the medical records, that there 'can be no proprietorship in information as information, because once imparted, belongs equally to both the patient and the health professional.

ethical duty not to use or disclose personal health information without their patient's consent unless a legal obligation or recognised exception exists.[62]

The *PCEHR Act* provides access by patients to their records uploaded onto the system, although ownership of the original record (statute or contract apart) is still presumably vested in the health provider who has created the record. The PCEHR Act is silent on the question of information ownership and property rights.

The *PCEHR Act* imposes limits on the collection, use and disclosure of health information included in the eHealth record. Pre-existing privacy laws at federal, state and territory levels also apply to protect information held about individuals, to allow them access to such information, as well as to determine whether it is accurate and collected for appropriate purposes.[63] Rights of access also exist under Freedom of Information legislation which provides patients with a statutory right of access to their personal health records held by public authorities.[64]

Although the ownership of the original eHealth record may vest in the health provider who has created it, the above example illustrates the many layers of rights, controls and obligations in relation to the information held in the eHealth record.

The *PCEHR Act* introduces new legal issues adding to the many existing issues as the records are held by a third party who operates the shared health record infrastructure (SEHR). There are questions as to whether the collation of records in a database by the SEHR conveys some proprietary rights (including copyright) to the infrastructure owner, over and above the copyright in the opinion provided by the health care provider (eg. doctor). There are also legal questions as to which organisation would be

---

[62] For a general discussion of the legal duty of confidentiality owed by medical practitioners, see LexisNexis, *Halsbury's Laws of Australia* (at 12 April 2011) [280-4000].

[63] *Privacy Act 1988* (Cth) extends the rights of access to documents held by private and Commonwealth public sector health providers. Also see *Health Records (Privacy and Access) Act 1997* (ACT), *Health Records and Information Privacy Act 2002* (NSW), *Health Records Act 2001* (NSW). There are no equivalent provisions in the other jurisdictions.

[64] *Freedom of Information Act 1982* (Cth) s 11. This legislation only has effect on agencies or organisations which are under the control of the Commonwealth; *Freedom of Information Act 1989* (ACT) s 10; *Government Information (Public Access) Act 2009* (NSW) s 3; *Right to Information Act 2009* (QLD) s 23; *Freedom of Information Act 1991* (SA) s 12; *Right to Information Act 2009* (TAS) ss 3, 7; *Freedom of Information Act 1982* (VIC) s 13; *Freedom of Information Act 1992* (WA) s 10.

responsible for satisfying freedom of information requests for information in the shared record.

Treating patient data as private property would hamper the creation of 'Big Data' datasets, even in an anonymised and de-identified form. It would be costly and laborious to secure rights to individual records as one of the benefits from 'Big Data' is the ability to combine, interrogate, mine and analyse large structured or unstructured, and multiple datasets to identify correlations to discover a cure for a disease, sickness or other information to improve medical knowledge.

We would have to wait and see whether our government, with the appropriate protective measures and safeguards to ensure confidentiality and privacy, would be open at a future period to allow data mining and analysis of the personally controlled electronic health record system.

5. CONCLUSION

What are the possible ownership constructs/ model for data in the era of 'Big Data'?

The White House Big Data Review advocates for increased clarity on the question of ownership and property rights in 'Big Data'. This matter should be high on our legislative agenda. However, we should tread with caution, as this chapter illustrates the layered complexities and issues pertaining to the granting of ownership rights in all data— especially data not currently embraced by our existing laws in relation to copyright, patent, confidential information and trade secrets. Defining 'ownership' in the context of 'Big Data' involves a balancing act between many vested interests including the interests of the 'purported' owner, the interests of competing third parties and the interests of the public to have access and use of the data.

From our example 2 in Section 4, it would appear that the framework of patient entitlements and protections afforded by the *PCEHR Act,* together with the myriads of privacy legislation (both state and federal) are remarkably similar to what patients

would enjoy if they own their eHealth information. The collective effect of these layered entitlements and protections may sometimes resemble the 'bundle of rights' in property law. This similarity suggests that property rights seem less appropriate and perhaps, 'should not be readily transferred and applied to more modern forms of wealth.'[65]

Has 'Big Data' Changed Our Perspective of Property Rights in Information? I believe it has. The average man or woman in our society would be surprised to discover that data in their possession, and in their computer or device, might not be owned by them in the traditional sense.

The *Criminal Code* recognises that the provider of the protected computer system has an inherent right to exclude unauthorised access to data, and is strikingly similar to what the provider of the protected computer system would have enjoyed if they own the data in the protected computer system.

In addition, Hemnes questions whether 'identity theft' laws have endowed some form of proprietary interests in personal information:

> If something is capable of being stolen, and if one can remedy the theft, does this not imply that it was owned in the first place?'[66]

We need a model for managing data. The White House Big Data Review suggests a model for the development of a set of tags to encode data:

> Tagging both enables precise access control and preserves links to source data and the purpose of its original collection. The end result is a taxonomy of rules governing where information goes and tracking where it came from and under what authority.[67]

---

[65] T C Grey, 'The Disintegration of Property' in J R Pennock and J W Chapman (eds), *Nomos XXII: Property* (New York University Press, 1980) 78. Cf S R Munzer, *A Theory of Property* (Cambridge University Press, 1990) 31–6.

[66] Hemnes, above n 22, 29.

[67] White House Big Data Review , above n 4, 28.

‘Big Data’ is created, derived and collected from a myriad of sources comprising ‘droplets’ or elements of data, some of which have the traditional intellectual property rights, some public domain, some licensed under different licensing schemes.

Rather than disrupting centuries of foundational work in the protection of intangibles work and creating more property rights in ‘drops of water’ which might gradually find their way to the greater ‘oceans of water’, the better approach and the challenge lies in understanding the many different contexts in which ‘Big Data’ comes into play. Enabling the legitimate and legal use of ‘Big Data’ datasets requires the securing of permission from the data custodians or owners of these disparate datasets of data, perhaps using the model for managing information above.

On a practical level, what should be the right balance and level of protection for data held in compilations, datasets and databases which may have been created at great expense without impeding the public’s access to and use of the facts and data for data mining and analysis in the era of ‘Big Data’? Getting the right balance is one of our biggest ultimate challenges moving forward.

We are being asked as a society to redefine a balance of private and public interests in data without any clear chart about where ‘Big Data’ will take us or even how ‘Big Data’ will work in years to come.

Currently, our legal framework falls short in providing clarity and certainty on the treatment and usage of data products in many areas of commercial endeavour. Law and policy makers, in reviewing and updating intellectual property and other laws, need to be aware of the specific challenges posed by ‘Big Data’ in the Digital Economy. Legal certainty on the ambit and scope of ownership in ‘Big Data’ are important to encourage further value and innovation.

With the pervasive use of computer technology, a rapidly growing percentage of our data is created automatically from the use of Internet of Things (IoT), sensors, devices, systems collecting transactional data and other devices. As most of the data collected comprise factually-based information, it is unlikely that they would be protected under our traditional intellectual property laws. Should rights be left to the realms of

contract, confidential information, trade secret, unfair competition laws and other mechanisms? Or should government provide the custodianship to enhance researchers' access to 'Big Data'?

Some might argue that the best way to encourage and reward creativity is to grant property rights to the creators and the right to control the assets they create. However, with such decentralised control, the cost of assembling the needed datasets for 'Big Data' mining and analysis would be exorbitant, requiring identification of the owners, coordination and interaction with owners of the required datasets.

The deployment of cloud technology and distributed networks are exacerbating the challenges in a rapidly globalised digital economy where data can easily traverse national borders. We need to work with our international counterparts to adopt international standards for the balanced protection of 'Big Data' and their fair use.

Unlike Australia, the EU has its *sui generis* database rights,[68] the UK has additionally implemented specific provisions to protect computer-generated works and in the US, the tort of misappropriation allows owners some control over the use that can be made of their 'Big Data' databases.

The time is ripe to act with some haste and reason in order for us to maximise and to leverage on the benefits of 'Big Data' while minimising its risks and improving our opportunities to foster and grow our digital economy, innovation and creativity. 'Big Data' is changing our world. Whether this is enough to evolve our laws in relation to the control and ownership of data remains to be seen. Ultimately, the question of whether property analogies are appropriate in the world of 'Big Data' would depend on the nature, context and characteristics of the data itself. 'Big Data' has and will continue to shape and change our perspectives of property rights in data.

[68] My research indicates that there have been some controversy and debate on the EU databases directive. However, due to word constraints in this paper, I have not ventured down the path to analyse the deficiencies of the current EU databases model. Readers should refer to Estelle Derclaye, *The Legal Protection of Databases: A Comparative Analysis* (Edward Elgar Publishing, 2008) 259 (Chapter 9), which examines the issues in greater detail.