Singapore Academy of Law
Law Reform Committee

Rethinking Database Rights and Data Ownership in an AI World

July 2020

SINGAPORE ACADEMY OF LAW

Singapore Academy of Law
Law Reform Committee

Rethinking Database Rights and Data Ownership in an AI World

July 2020

Part of the *Impact of Robotics and Artificial Intelligence on the Law* series

Members of the Robotics and Artificial Intelligence Subcommittee

1. The Honourable Justice Kannan Ramesh (co-chair)
2. Charles Lim Aeng Cheng (co-chair)
3. Chen Siyuan
4. Desmond Chew
5. Josh Lee Kok Thong
6. Gilbert Leong
7. Beverly Lim
8. Sampson Lim
9. Ronald Wong Jian Jie
10. Yvonne Tan Peck Hong
11. Yeong Zee Kin

The report was edited by Simon Constantine, Deputy Research Director, Singapore Academy of Law.

An electronic copy may be accessed from the Singapore Academy of Law website *https://www.sal.org.sg/Resources-Tools/Law-Reform/Law-Reform-e-Archive*.

National Library Board, Singapore Cataloguing in Publication Data

Name(s): Singapore Academy of Law. Law Reform Committee. | Constantine, Simon, editor.
Title: Rethinking database rights and data ownership in an AI world / Singapore Academy of Law, Law Reform Committee ; edited by Simon Constantine.
Other title(s): Impact of robotics and artificial intelligence on the law
Description: Singapore: Law Reform Committee, Singapore Academy of Law, [2020]
Identifier(s): OCN 1158456322 | ISBN 978-981-14-6597-0 (paperback) | ISBN 978-981-14-6598-7 (ebook)
Subject(s): LCSH: Databases–Law and legislation–Singapore. | Big data–Law and legislation–Singapore.
Classification: DDC 343.59570999–dc23

ISBN 978-981-14-6597-0 (softcover)
978-981-14-6598-7 (e-book)

About the Law Reform Committee

The Law Reform Committee ("LRC") of the Singapore Academy of Law makes recommendations to the authorities on the need for legislation in any particular area or subject of the law. In addition, the Committee reviews any legislation before Parliament and makes recommendations for amendments to legislation (if any) and for carrying out law reform.

Comments and feedback on this report should be addressed to:

Law Reform Committee
Attn: Law Reform Director
Singapore Academy of Law
1 Coleman Street
#08-06 The Adelphi
Singapore 179803
Tel: +65 6332 4070
Fax: +65 6333 9747
Email: *lawreform@sal.org.sg*

IMPACT OF ROBOTICS AND ARTIFICIAL INTELLIGENCE ON THE LAW

SERIES PREFACE

It has been said that we are at an inflection point in the development and use of Artificial Intelligence (AI). The exponential growth in data in the past decade – from 2 trillion gigabytes in 2010 to around 33 trillion at the end of 2018, and an anticipated 175 trillion by 2025 – has enabled giant datasets to be compiled and used as the basis for developing ever-more sophisticated AI systems.

Those systems are in turn being used – in commercial, military, consumer and other contexts – to enhance humans' ability to carry out tasks, or to replace humans altogether. From self-driving cars and robotic carers, to autonomous weapons and automated financial trading systems, robotic and other data-driven AI systems are increasingly becoming the cornerstones of our economies and our daily lives. Increased automation promises significant societal benefits. Yet as ever more processes are carried out without the involvement of a 'human actor', the focus turns to how those robots and other autonomous systems operate, how they 'learn', and the data on which they base their decisions to act.

Even in Singapore, which ranked first in the 2019 International Development Research Centre's Government Artificial Intelligence Readiness Index, questions inevitably arise as to whether existing systems of law, regulation and wider public policy remain 'fit for purpose', given the pace and ceaselessness of change. That is, do they encourage and enable innovation, economic growth and public welfare, while at the same time offering protection against misuse and physical, financial or psychological harm to individuals?

To this end, the Singapore Academy of Law's Law Reform Committee ('LRC') established a Subcommittee on Robotics and Artificial Intelligence to consider, and make recommendations regarding, the application of the law to AI systems.
Having considered current Singapore law, as well as legal and policy developments in other parts of the world, the LRC is now publishing a series of reports addressing discrete legal issues arising in an AI context.

There is currently much work being undertaken at a national and international level in this field. Domestically, the Singapore Government has published the second edition of its Model AI Governance Framework and launched a National Artificial Intelligence Strategy to reap the benefits of systematic and extensive application of new technologies. The LRC hopes that its reports will complement and contribute to these efforts and help Singapore law – through legislation or 'soft law' – to develop in a way that fosters socially and economically beneficial development and use of robotic and AI-driven technologies.

The series does not purport to offer comprehensive solutions to the many issues raised. The LRC hopes, however, that it will stimulate systematic thought and debate on these issues by policy makers, legislators, industry, the legal profession and the public.

OTHER REPORTS IN THIS SERIES

- Applying Ethical Principles for Artificial Intelligence and Autonomous Systems in Regulatory Reform (published July 2020)
- Report on the Application of Criminal Law to the Operation of Artificial Intelligence Systems and Technologies (*forthcoming*)
- Report on the Attribution of Civil Liability for Accidents Involving Automated Cars (*forthcoming*)

TABLE OF CONTENTS

SERIES PREFACE ..... **iv**

EXECUTIVE SUMMARY **1**

CHAPTER 1 INTRODUCTION **5**

CHAPTER 2 UNLOCKING THE VALUE OF DATABASES **7**

A Background: database creation 7
1 The typical data flow process 8
2 Database design 9
B Legal mechanisms for protecting databases 10
1 The current position in Singapore 10
(a) Copyright 10
(b) Patents 15
2 Approaches in other jurisdictions to protecting databases 16
(a) *Sui generis* database rights 16
(b) United States' reluctance to adopt the SGDR 20
3 Other potential models to balance the interests of database owners and users 22
(a) Competition law 22
(b) Data portability 23
(c) Tort of unfair competition 24
(d) Contract 25
C Recommendations 25
1 No real impetus for Singapore to introduce a sui generis right 25
2 Re-examine the fundamentals of authorship under Singapore's Copyright Act 29
3 Concluding observations 31

CHAPTER 3 DATA OWNERSHIP **32**

A Classifications of data 33
1 Non-personal data 33
2 Personal data 36
B Merits of granting property rights over *personal* data 37
1 The existing legal framework 38
(a) Copyright, confidentiality and privacy 38
(b) Data Protection and incidents of ownership 40
2 Challenges with conferring ownership rights over personal data 45
C Recommendations 47

CHAPTER 4 CONCLUSION **49**

GLOSSARY **51**

RETHINKING DATABASE RIGHTS AND DATA OWNERSHIP IN AN AI WORLD

EXECUTIVE SUMMARY

1 This Report of the Singapore Academy of Law's Law Reform Committee ('LRC') Subcommittee on Robotics and Artificial Intelligence considers the legal issues regarding:

a. who controls or has rights over the 'big data' databases that underpin many new and emerging Artificial Intelligence (AI) technologies, and

b. how best to ensure that those who contribute data to those databases retain an appropriate degree of control over and access to that data.

2 The 'Big Data' revolution of the past decade has seen huge growth in the generation and collection of personal and non-personal digital data. This in turn has driven the creation of large datasets and databases to aggregate and arrange that data.

3 Those datasets and databases are the 'stock feed' on which AI systems – and machine learning AI systems in particular – rely. As such, the legal and policy landscape governing those databases is of crucial importance. Legal and policy shortcomings at the 'database' level will have ripple effects (potentially significantly amplified) at the 'application' level.

Unlocking the Value of Databases

4 The creation of 'big data' databases has the potential to drive significant societal benefits. Such benefits are likely to be maximised where third parties are able to gain access to databases and to combine the data in them with other datasets in their possession to derive higher quality data analysis.

5 From a public policy perspective, therefore, a balance must be struck between:

a. on the one hand, rewarding both creators of original works and those that put time and effort into ensuring data quality, so as to encourage further investment and innovation; and,

b. on the other, ensuring sufficient access for others to those databases, so as to maximise their productive uses and benefits, drive competition and enable follow-on innovation.

6 Today, database protection in Singapore comes primarily from the patent and copyright regime. However, such protection is limited to

elements that meet the requisite level of originality. Electronic databases are compilations of facts and information, and as such do not necessarily sit comfortably within these intellectual property law regimes, as they are currently formulated.

7 Specifically, for compilations to attract copyright, both a) "application of intellectual effort, creativity or the exercise of mental labour, skill or judgment" by a human author, and b) some element of selection or arrangement of the data, must be demonstrated. This creates challenges when applied to databases, which may, for example, be machine-generated, be compiled systematically rather than 'creatively', or involve some process of 'selection' only *after* the compilation has been created.

8 Similarly, patent laws may provide only partial protection (of the particular software in question, rather than of the database as a collection of data or its structures). Indeed, patent protection may be wholly unavailable where databases are built using open-source or collaborative tools, as is increasingly common.

9 One possible response to such limitations would be the creation of a standalone '*sui generis*' database right, similar to that which exists in the European Union ('**EU**'). While such a *sui generis* right would appear to provide more comprehensive (or at least more certain) protection for all stages of the creation of databases, the EU experience suggests that it has done little to spur investment in the creation of new databases. By contrast, the United States has no such *sui generis* right, yet has seen far stronger growth in database creation.

10 Other mechanisms also exist in Singapore law to balance the interests of creators with the benefits of enabling access for legitimate uses. These include competition law, data portability, torts of unfair competition and contract law. Taken together, such laws may provide some further stimulus to both the beneficial production and use of databases.

11 In light of the above:

a. We do not recommend creation of a *sui generis* database right in Singapore, given i) the limited evidence of its effectiveness in driving database production, and ii) crucially, the broader jurisprudential differences between the EU and Singapore intellectual property regimes.

b. Instead, we recommend that:

   i. copyright protection of computer-generated works be recognised through legislative amendments modelled on equivalent UK copyright laws, with guidance provided in the interim on when computer generated works enjoy copyright protection; and

ii. further clarity is given (for example through subsidiary legislation, or through administrative guidance or other 'soft law' measures) as to: a) how compilation rights apply for the copyright protection of electronic databases, and b) how records of authorship of databases can be properly maintained.

Data Ownership

12 In the face of mass collection of data, in particular from 'smart' devices, and given the sensitivity and potential commercial value of such data, there have been calls for formal property rights to be accorded to data.

13 Presently, for *personal data*, the Personal Data Protection Act ('**PDPA**') grants individuals certain rights over their personal data, but does not confer legal ownership (in the way that, say, intellectual property rights do over creative works). Similarly, while privacy or confidentiality laws may provide protections for data subjects in specific circumstances, such protections are not founded in notions of 'property' or 'ownership', nor do they claim to create such rights.

14 For *non-personal* data, either the law of confidence or various sectoral legislations may offer protection to a person or entity that creates or controls data (above and beyond any technical or contractual protections they may themselves apply). However, issues can still arise if, for example, the person or entity whose activities generated the data cannot themselves access it. That may include risks of them being 'locked-in' to a particular data processor or hardware provider, or possible wider limitations on competition and innovation.

15 Given the nature of data, there are fundamental difficulties – on grounds of jurisprudential principle and policy – to using ownership and property rights as legal frameworks to control data. In addition, any attempt to create such a right would mean significant disruption to established legal frameworks.

16 As such, our conclusions are as follows:

a. For *personal data*, we consider that existing and incoming data protection laws currently provide data subjects with the crucial ability to exercise adequate control over their data.

b. For *non-personal* data, the issues are to some extent less acute. However, we see merit in consideration being given to whether a right akin to data portability should also be introduced for such non-personal data. Laws promoting the transfer of non-personal data in both the EU and Australia provide possible precedents for such a right. Similarly, the planned introduction of data portability obligations for personal data under the PDPA offers a chance to assess the effectiveness of such rights and to consider their possible extension to non-personal data.

17 It is hoped that the analysis and recommendations in this report will help facilitate the development of Singapore law in ways that help drive the benefits offered by data and stimulate innovation, while protecting the important rights and liberties of individuals.

CHAPTER 1

INTRODUCTION

1.1 The Fourth Industrial Revolution is one that is characterised by the convergence of emerging technologies (such as artificial intelligence ('**AI**'), Internet of Things ('**IoT**') technologies and robotics) and their "interaction across the physical, digital and biological domains".[1]

1.2 At the heart of the Fourth Industrial Revolution is "big data" – vast, rapidly-growing collections of data of different types, from which detailed insights can be derived using advanced technology and analytical methods.[2] Such big data has the potential to be transformative and unlock enormous opportunities for Singapore[3] (and, indeed, to give rise to significant policy and other challenges).

1.3 Advances in technology have allowed for the near-instantaneous collection of raw data – both personal and non-personal – from various devices (for example, smart home speakers and smart watches), forming datasets that can then be processed and aggregated into databases.[4] However, the utility of big data is only truly realised when those databases are mined to create information that is useful for end users, and from which they can extract value[5] – for example, enabling public transport operators to spot maintenance issues,[6] or e-commerce websites to tailor product recommendations to customers.

1.4 Seen in that context, datasets and databases are the 'stock feed' on which AI systems rely, especially those AI systems that utilise machine learning. The legal and policy landscape governing those databases is

---

1 Klaus Schwab, *The Fourth Industrial Revolution* (Geneva: World Economic Forum, 2016) at 8.

2 We note that definitions of 'big data' vary. Broadly, however, they are united by their reference to the volume, velocity and variety of the data in question.

3 See Committee on Future Economy, *Report of the Committee of the Future Economy* (7 February 2017) at [37]. The Committee on Future Economy has identified data to be an "increasingly important source of comparative advantage" and that greater ability to use it productively was needed in order to propel Singapore forward.

4 E Douilhet and Argyro Karanasiou, "Legal Responses to the Commodification of Personal Data in the Era of Big Data: The Paradigm Shift from Data Protection towards Data Ownership" in Manoj Kumar Singh and Dileep Kumar G (eds), *Effective Big Data Management and Opportunities for Implementation* (Hershey, Pa: Information Science Reference, 2016) at 130–139.

5 *Ibid.*

6 For example, the value and power of big data was demonstrated when Singapore government agencies used it to resolve mysterious intermittent signal interference from a rogue train on the Circle Line: *Rogue Train: A Big Data Story*, GovTech Singapore (28 December 2016) <*https://www.tech.gov.sg/media/technews/rogue-train-a-big-data-story*> (accessed 10 June 2020).

therefore of crucial importance: legal and policy shortcomings at the 'data' level will have ripple effects (potentially significantly amplified) at the 'application' level.

1.5 For example, poor practices in data governance may result in data quality problems (e.g. a dataset ends up not being representative of the intended consumer market). Such quality issues can then result in unexpected product behaviour (e.g. unintended discrimination or biased predictions) and, in turn, damage a company's commercial reputation. The challenge for policy makers is to design laws and policies that enable the full potential of big data to be realised and encourage innovation, while protecting the personal rights and wellbeing of citizens.

1.6 There are many different legal and policy issues surrounding the collection and use of data (in particular personal data). They include issues around privacy and data sharing, data quality (including errors and biases), competition and access, cybersecurity and more. Many of these issues are beyond the scope of this report or have been addressed by others.[7] Instead, we focus on two issues that – even if not the only issues at play – are of fundamental importance:

a) who controls or has rights over such 'big data' databases, and

b) how best to ensure that those who contribute data to those databases retain an appropriate degree of control over, and access, to that data.

1.7 We believe that having the right legal framework for analysing these issues around ownership and control can also provide a launchpad for solving other, downstream issues regarding datasets and data – issues such as what responsibilities or duties should be imposed for ensuring the quality of input data used for training or operating AI machine learning models, or who should hold any rights in the derivative work that is the output of such processes.

7 See, for example, in relation to data quality, Personal Data Protection Commission of Singapore, *Model AI Governance Framework (Second Edition)*, <*https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf*> (accessed 10 June 2020); and in relation to issues surrounding the use of data for model training, Yeong Zee Kin, "Legal Issues in AI Deployment", Law Gazette (February 2019) <*https://lawgazette.com.sg/feature/legal-issues-in-ai-deployment/*> (accessed 10 June 2020).

CHAPTER 2

UNLOCKING THE VALUE OF DATABASES

A BACKGROUND: DATABASE CREATION

2.1 Given the significant value of big data, it is little wonder that governments and organisations are investing heavily in the acquisition and creation of electronic databases. Moreover, the network effect of "connected users, devices and sensors"[8] means that – while acquiring, generating and obtaining data has never been easier – there is a risk that big data is concentrated in the hands of a small handful of actors in the market.

2.2 From a public policy perspective, the societal benefits of big data are maximised when third parties are able to gain access to the databases, allowing them to combine the data in them with other datasets in their possession to derive higher quality data analysis.[9] However, organisations may have little financial incentive to invest in databases given the ease of a third party in reproducing its contents.[10]

2.3 This raises tensions between three drivers, which may often pull in different directions: (1) rewarding creators of original works to encourage further innovation and investments into AI and big data; (2) recognising the investment of skill, time and effort in data governance in order to ensure quality data;[11] and (3) ensuring sufficient access by third parties to maximise the benefits of the original works.[12]

---

8 Competition Commission of Singapore in collaboration with the Intellectual Property Office of Singapore and the Personal Data Protection Commission, Singapore, *Data: Engine for Growth – Implications for Competition Law, Personal Data Protection and Intellectual Property Rights* (Singapore: Competition Commission of Singapore, 2017) at [16] <*https://www.cccs.gov.sg/-/media/custom/ccs/files/media-and-publications/publications/occasional-paper/ccs-big-data-paper-16-aug-2017nonconfi-final.pdf*> (accessed 10 June 2020). ('**Data: Engine for Growth**')

9 Yip Man, "Protecting Consumer's Personal Data in the Digital World – Challenges and Changes" [2018] Personal Data Protection Digest 104 at [3] <*https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4862&context=sol_research*> (accessed 10 June 2020).

10 Daryl Lim, "Re-defining the Rights and Responsibilities of Database Owners Under Competition Law" (2006) 18 Sing Acad LJ 418 at [4].

11 See Personal Data Protection Commission of Singapore, *Model AI Governance Framework (Second Edition)*, above, n 7 at 38.

12 Paul Chan, "Distributing the Economic Benefits of Databases: New Wine, New Bottles" in Lee Seiu Kin (ed), *Global Technology Law Conference 2015: The Future of Money and Data* (Singapore: Academy Publishing, 2016) at [5], and Tan Tee Jim, "New Law for Compilations and Databases in Singapore?" (2012) 24 Sing Acad LJ 745 at [5].

2.4 This tension is not a new one; it has been discussed in the context of compilations and non-electronic databases such as telephone books.[13] However, it takes on greater urgency in the context of big data, given the scale and value of electronic databases. Providing a sensible legal framework for resolving these tensions will help to achieve the policy objectives of enabling the use of data to support innovation and competition for the ultimate benefit of consumers.

2.5 Much of the discussion in relation to how databases and compilations should be protected, and how the competing objectives above can be balanced, has centred on the adequacy (or otherwise) of Singapore's intellectual property law regime. As will be seen, these laws currently remain the primary means of affording databases protection, notwithstanding their somewhat ill-fit with databases as compilations of data and facts. Accordingly, this report will examine the state of database protection under Singapore's existing intellectual property laws, and consider the possible directions that Singapore can take to ensure that databases are protected in a manner that best balances the societal goals outlined above. In particular, it will evaluate the model of a *sui generis* database right (i.e. a standalone database right), which exists in the European Union, but has been clearly rejected in the United States.

2.6 In addressing the legal issues associated with databases, it is useful briefly to explain at the outset: (1) databases' place and role in the typical flow of data from its raw form to ultimately being capable of interpretation and use by end users; and (2) the technological set up of databases that process big data.

1 The typical data flow process

2.7 There are usually four key stages in the life cycle of a data flow, namely (1) collection; (2) integration; (3) mining; and (4) usage:[14]

1) At the first stage, an organisation may collect or obtain raw data by various means, such as the direct collection of raw data from IoT devices, or the buying or obtaining of data from third parties (for example, publicly available data at *data.gov.sg*). These data may be collected directly or indirectly from the different source points, and may include personal data (for example, biometric data).[15]

13 Tan, *ibid.*

14 *Data: Engine for Growth*, above, n 8 at [22]; and Douilhet and Karanasiou, above, n 4 at 130–139.

15 The collection of biometric data may engender privacy issues. For further information on this issue, see Gilbert Leong, Foo Maw Jiun and Desmond Chew, "Regulation of Biometric Data under the Personal Data Protection Act" [2018] Personal Data Protection Digest 134.

2) Thereafter, the data collected from the various source points will be combined, re-formatted to form consistent datasets, and eventually integrated into a database.[16]

3) This enables an organisation to mine the data (often using predictive modelling tools or other analytical software) to create useful information for the end users.

4) Lastly, the information will be reviewed by the organisation or the end users and exploited through various means.

2.8 As is evident from the cycle of a data flow process, databases play an important role in: (1) the collection of huge amounts of raw data; (2) the investment of time and effort in ensuring data quality so that meaningful insights may be obtained on analysis; and (3) the processing, combining and standardising of datasets so that they may be analysed and useful (and usable) insights derived from them.

2 Database design

2.9 There are two aspects to electronic databases that are of importance: (1) the storage functionality of the databases; and (2) the display of the data or information in the databases to the end users.

2.10 Data in electronic databases are typically stored in tables. The effectiveness of a database design (or schema) lies in the structure of the tables, and how relationships are built between the tables. The relationships between tables, when done well, can store large amounts of data efficiently.[17] However, the relationship and structure of databases are typically abstract and are not visible to the end users.

2.11 The data stored in the databases may be structured, semi-structured and/or unstructured.[18] With the growth of big data,[19] it is increasingly possible for databases of unstructured data (such as photographs and audio recordings) to be accessed and manipulated without the need for metadata to be extracted and stored in structured form.

16 *Data: Engine for Growth*, above, n 8 at [22].

17 Kim Nguyen, "Relational Database Schema Design Overview", *Medium* (4 October 2017) <*https://medium.com/@kimtnguyen/relational-database-schema-design-overview-70e447ff66f9*> (accessed 10 June 2020).

18 Broadly speaking, 'structured' data is data that is highly-organised and formatted according to pre-defined fields, making it simpler to search and analyse. Unstructured data, by contrast, is not organised or formatted in any pre-defined way, and thus is harder to interrogate. See Bernard Marr, "What's The Difference Between Structured, Semi-Structured And Unstructured Data?" *Forbes* (18 October 2019) <*https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data*> (accessed 10 June 2020).

19 Specifically, database technologies with enhanced capacity to index and query unstructured data, and thus to handle large files, such as visual media.

2.12 End users do not see the complexity of the databases. Instead, what data is displayed, and how, depends on the function and purpose of the particular user interface or display. Such display of databases depends on the software or application that, in response to a user's query, selects and processes the appropriate data from the database into a comprehensible output. The extent of useful information that can be derived from databases is therefore tied to the functionalities, purposes, capabilities and sophistication of the retrieval software or application.

B LEGAL MECHANISMS FOR PROTECTING DATABASES

1 The current position in Singapore

2.13 The present landscape of database protection in Singapore is primarily governed by the patent and copyright regime.[20] Even then, however, those regimes are only partially applicable: electronic databases are compilations of facts and information, which do not necessarily sit comfortably within the traditional intellectual property law frameworks.[21] Section 7A of the Copyright Act[22], for example, expressly states that literary works protected by copyright law include "compilation in any form". While this might suggest databases are accorded due protection, the Court of Appeal in *Global Yellow Pages v Promedia Directories Pte Ltd*[23] ("***Global Yellow Pages***") stated that – notwithstanding that a third party may appropriate "data or facts that represents the fruit of an investment" – this is "simply not within the purview of copyright law".[24]

2.14 This section will explore how the patent and copyright regimes struggle to cope with databases and the rise of big data, before exploring the experiences of other jurisdictions in addressing database protection and various alternatives that have been proposed to address the perceived deficiencies of database protection in Singapore.

(a) Copyright

2.15 Article 10(2) of the Agreement on Trade-Related Aspects of Intellectual Property Rights ('**TRIPS Agreement**') requires signatory members to protect "compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations". To comply with Article 10(2) of the TRIPS Agreement, Singapore protects databases

20 While some protections may be available to database owners and creators through the law of contract, as will be discussed further below, this too provides only narrow and/or incomplete protection.
21 Chan, "Distributing the Economic Benefits of Databases", above, n 12 at [6].
22 Cap 63, 2006 Rev Ed.
23 [2017] 2 SLR 185, CA.
24 *Global Yellow Pages*, above, n 23 at [34].

primarily in the form of section 7A of the Copyright Act,[25] which expressly recognises compilation as a form of literary work.

2.16 Insofar as humans design electronic database software, it is relatively uncontroversial that such software as a whole can be capable of copyright protection on the basis that it is a literary work, subject to meeting the relevant criteria under the Copyright Act.[26]

2.17 However, there is a fundamental difficulty in using the copyright regime to protect the *processes and output* of electronic databases. That is because copyright law is based on the notion that a human author expending intellectual effort must be involved in the compilation of the databases in order to enjoy copyright protection. In *Global Yellow Pages*, the Court of Appeal commented that:[27]

> [...] for copyright to subsist in any literary work, there must be an *authorial* creation that is *causally connected* with the *engagement of the human intellect*. By the human intellect, we mean the ***application of intellectual effort, creativity or the exercise of mental labour, skill or judgment***. Effort (even intellectual) that is applied *not* towards the authorial creation but towards other ends such as the verification of facts will not be relevant in this context even if such verified facts might be the eventual subject of the authorial creation.
>
> [Emphasis in bold italics added.]

2.18 The reality is that it is typically difficult to attribute a database's creation to a particular human author. Companies, for example, often design large electronic databases collaboratively due to the vast amount of datasets that are collected. However, Singapore's copyright law assumes that there must be a natural person[28] to whom the copyright in a work can be given, which does not fit well with the way in which electronic databases are created:[29]

- Take, for example, location data gathered via GPS sensors on a person's phone and automatically sent to a mapping app developer's servers. That person's (human) activity generated the data, but they cannot be said to have consciously authored it. Neither have the app developer's staff had any role in authorship of the data in the copyright sense.

---

25 Cap 63, 2006 Rev Ed.

26 Ng-Loy Wee Loon, *The Law of Intellectual Property of Singapore* (2nd Ed) (Singapore: Sweet & Maxwell Asia, 2014) at [6.1.1] and [6.1.15].

27 *Global Yellow Pages*, above, n 23 at [24] (emphasis in the original, except for the text in bold italics).

28 Section 27 of the Copyright Act states, broadly, that copyright subsists in work authored by a "qualified person", defined as "a citizen of Singapore or a person resident in Singapore".

29 *Asia Pacific Publishing Pte Ltd v Pioneers & Leaders (Publishers) Pte Ltd* [2011] 4 SLR 381 at [72], CA.

- Alternatively, consider raw-machine generated databases (for example, manufacturing equipment may generate data about its throughput). In this context, the selection of appropriate data to be accessed and analysed by sophisticated analytics engines is often done without human intervention. In the absence of any human author, most jurisdictions deem raw-machine generated databases not to be deserving of any protection due to the lack of "an intellectual effort and/or [...] any degree of originality".[30]

2.19 Further, a second difficulty arises in applying the copyright law concepts of selection and/or arrangement of the data to electronic databases:

- On the one hand, one might argue that there is a degree of selection at the design phase, when the software designer decides *what categories* of data (for example, personal particulars, location data and step count) will be collected, *how* (for example, user-input through online forms, or IoT sensors on smart devices or manufacturing equipment) and *when* (for example, when there is an error detected).
- However, once that application system is designed, deployed and used in a production environment, data is automatically collected continuously throughout its operation, often in an indiscriminate and unorganised manner.[31] This 'selection' process (with minimal to no human involvement) is not visible and may not even be apparent to the end users.

2.20 How electronic databases store data is different from how records from these electronic databases are presented and displayed to users. In the typical three-tier system architecture,[32] data stored in the 'data layer' (i.e. the electronic database) is selected by the 'application' (or 'logic') layer (e.g. user-provided input parameters are used to perform a query from the database) and displayed to the user in the 'presentation layer' (e.g. the report showing records from the database that match the parameters used to query the database).

30 *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: "Building a European Data Economy"* (COM/2017/09 final) (Brussels, European Commission, 2017) at 10, section 3.3 <*https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017DC0009&from=EN*> (accessed 10 June 2020).

31 Mark Vincent and Katrina Crooks, *Australia: Can a Database be Protected by Copyright?,* Mondaq (7 February 2014) <*http://www.mondaq.com/australia/x/290668/Copyright/Can+a+database+be+protected+by+copyright*> (accessed 10 June 2020).

32 See, for example, 'Three Tier Architecture' in *Multitier Architecture* <*https://en.wikipedia.org/wiki/Multitier_architecture#Three-tier_architecture*> (accessed 10 June 2020).

2.21 It is questionable whether this fits into the traditional copyright understanding of 'compilations', which was intended to protect compilations of, for example, verses, poems or even music albums and movie collections.

2.22 The necessary elements of selection take place at the logic layer and an arrangement is displayed at the presentation layer, but these are divorced from the data layer. Crucially, these activities take place *after* the individual records forming the database have been compiled. Thus, it is simplistic – and often probably wrong – to look at the software application that is displayed on screen and conclude that copyright subsists. It is necessary to look 'under the hood' and understand the data collation process, so as to analyse these issues accurately.

2.23 For electronic databases that have a data-entry form that is filled in by a human user, data is typically entered one record at a time and thereafter stored *sequentially* in the electronic database. In this context, it cannot be said that there is an arrangement of data. Likewise, where data is collected through other input sources (for example, IoT sensors), it cannot be said that any arrangement of data takes place.

2.24 Nor is it clear that copyright protection would subsist in the display of data to the user at the presentation level. When information from electronic databases is retrieved and displayed on a screen to a user, the information displayed is retrieved from the electronic database by application codes, which operate based on the user's input parameters. For example, the user's query may be to see all records where app users living in Ang Mo Kio had step counts above 10,000 and for those to be arranged by their postal code, for the purpose of sending them supermarket discount vouchers. The selection and arrangement take place on retrieval and are executed by code. It will be difficult to claim that this extract from the electronic database is a compilation, or that there is sufficient originality in the selection and arrangement.[33]

2.25 Moreover for copyright to subsist, originality has to be found in the selection and arrangement *at the point of time* of creation of the compilation. What is presented on the screen is merely an extract of the underlying electronic database. On this analysis, the relevant point in time for determining whether copyright subsists is at the stage of data-entry. As discussed in the preceding paragraphs, the dynamics of sequential data-entry call into question whether there is any selection or arrangement.[34] These criticisms may make it difficult to sustain an argument that there was

33 Paula Baron, "Back to the Future: Learning from the Past in the Database Debate" (2001) 62 Ohio State LJ 879 at 900 – 901.

34 While it might be argued that the operative business rules may be sufficient 'selection', the fact that records are entered into the database sequentially calls into question whether there was any 'arrangement'.

sufficient originality at the time a dataset was created to warrant protection as a compilation under copyright law.

2.26 As the foregoing analysis illustrates, there are serious issues with the copyright protection of databases that have not thus far been fully ventilated in the cases that have proceeded before the Singapore courts. In summary:

- The skill and effort that copyright protects in the selection and arrangement of compilations does not map across easily to the way modern electronic databases are designed and how data is collected.
- While there may be some selection of the categories of data to be captured and stored in the electronic database, there is no conscious selection once the application system is operational, as all records meeting the pre-determined selection criteria are collected.
- Nor is there any arrangement in a) the electronic database (as this simply stores all records sequentially), or b) the display of data on a screen in line with the user's requested parameters (as this takes place *after* the database and its records have been created).

2.27 Furthermore, insofar as the Court of Appeal has addressed copyright in databases, its re-emphasis of the centrality of creativity makes clear that only truly original aspects of selection and arrangements in databases are appropriate for protection.[35] In reality, however, the most valuable databases (and thus those arguably most deserving of protection) are specifically designed to be systematic and comprehensive, and do not necessarily involve creative effort or human intellect.[36]

2.28 The above paragraphs show the difficulty of applying copyright concepts to electronic databases. However, this may be because copyright laws are not the appropriate laws to address the unique technological nature of databases in the first place. Article 9(2) of the TRIPS Agreement provides that copyright protection is meant to extend to "expressions, and not to ideas, procedures, methods of operation or mathematical concepts". Thus, copyright was not intended to cover methods of assembly, rules and procedures in database designs or even the data contained within the databases.[37] Even if copyright attaches to a database as a compilation, any such protection will necessarily be weak, and fail to protect other important, valuable aspects of the database, for example, the investment of

35 This is the case in the United States as well (see Baron, "Back to the Future: Learning from the Past in the Database Debate", above, n 33).

36 Chan, "Distributing the Economic Benefits of Databases", above, n 12 at [27].

37 *Id.* at [24]; and Lim, "Re-defining the Rights and Responsibilities of Database Owners under Competition Law", above, n 10 at [15].

routine and often 'grunt' effort in data-entry and the disciplines necessary for assuring data quality.

2.29 The difficulties discussed are not insurmountable. Copyright law has proved historically to be most malleable when tasked to protect new technologies. It has, for example, been extended to protect computer programs as literary works and older forms of databases as compilations. But it will be necessary for the issues to be addressed *directly*. As such, this can be an area where we consider that either IPOS administrative guidance (or similar 'soft law' measures within its remit) or subsidiary legislation can be provided to shape the interpretation and application of copyright principles to the protection of electronic databases.

(b) Patents

2.30 It is also possible to seek protection of aspects of electronic databases in the form of patents, where they meet the criteria under the Patents Act.[38]

2.31 In the traditional bespoke or commercial-off-the-shelf ('**COTS**') database software model, organisations would design, develop and customise software that is capable of enabling "users to define, create, maintain and control access to the database".[39] Such inventions are capable of protection under the patent regime in the form of a software patent, with claims to protect the schema, the structure of tables and the relationships between tables.

2.32 However, the scope of such patents is usually limited to the particular software in question (i.e. the database management system, or DBMS), and do not extend to the protection of the database as a collection of data, or to its structures. Further, patent offices remain cautious to set boundaries for issuing database-related patents.[40] Moreover, with the rise of big data, software designers are increasingly likely (or required) to use open source tools and a cloud hosting environment in order to design their electronic databases.[41] Such reliance on structures, datasets and information that are already publicly available (and therefore lack 'novelty'

38 Cap 221, 2005 Rev Ed.

39 Thomas M Connolly and Carolyn E Begg, *Database Systems: A Practical Approach to Design Implementation and Management* (6th ed) (Harlow, Essex: Pearson, 2015) at 64.

40 The European Patent Office ('**EPO**'), for example, has issued guidelines to emphasise that the classification of abstract data records without indication of a technical use of the resulting classification is not per se a 'technical' purpose. *Guidelines for Examination*, European Patent Office (17 September 2018), section 3.3.1 ("Artificial Intelligence and Machine Learning") <*https://www.epo.org/law-practice/legal-texts/html/guidelines2018/e/g_ii_3_3_1.htm*> (accessed 10 June 2020).

41 Joel E Lehrer, "United States: Patenting 'Big Data'", Mondaq (20 August 2012) <*http://www.mondaq.com/unitedstates/x/192640/Patent/Patenting+Big+Data*> (accessed 10 June 2020).

under patent law), may further limit the degree to which patent protection is available.[42]

2.33 That is not to say that aspects of the DBMS are incapable of being patented. With the advance of AI tools, it is conceivable that software designers may explore new and innovative methodologies of designing DBMSs (for example, non-relational databases), as well as processes relating to the use of such databases[43] – all of which could, in principle, be patentable.

2.34 In summary, therefore, while patents do play a role in protecting aspects of DBMSs, the fact that the underlying data, datasets and/or structures that comprise the databases may not be protected by patent law materially constrains the extent and utility of protection available to database owners.

2 Approaches in other jurisdictions to protecting databases

2.35 There have been various attempts by jurisdictions, including the European Union (‘**EU**’) and the United States, to employ different legal devices to protect databases. The most experimental one to date is that of the standalone database rights implemented by the EU. As will be seen, however, that ‘experiment’ has not been without its challenges.

(a) *Sui generis* database rights

2.36 In a *sui generis* database right (‘**SGDR**’) model, the database itself – not the underlying data – enjoys the protection offered by property rights. The only legislative instance of a successful implementation of SGDR is the EU’s enactment of the Database Directive 1996 (‘**DD**’).[44] The Directive provides for two-tiers of database protection:

- (a) Copyright protects the structure of the database (Article 3); and
- (b) the SGDR accords secondary protection against the “extraction and/or re-utilisation of the whole or a substantial part” of the database content (Article 7).

2.37 Since the SGDR can subsist regardless of whether the database or its contents is a copyright work (Article 7(4)), it thus fills the gap between the

42 Chan, “Distributing the Economic Benefits of Databases”, above, n 12 at 200.

43 In the United States, for example, patents have been granted with respect to novel querying techniques of databases: *Modelling and Implementing Complex Data Access Operations Based on Lower Level Traditional Operations*, United States Patent No US7,949,685 B2. See also Lehrer, “Patenting ‘Big Data’”, above, n 41.

44 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases, OJ L 77 27.03.96, p. 20. <https://eur-lex.europa.eu/eli/dir/1996/9/oj> (accessed 10 June 2020).

need for database protection and the deficiency of protection that copyright offers (as a result of the relatively high threshold of originality that EU copyright law in particular requires for works to be eligible for copyright protection).[45] By contrast, all that is necessary for the SGDR to subsist is that the database owner shows a "qualitatively and/or quantitatively […] substantial investment" in the "obtaining, verification or presentation" of the content (Article 7(1)).

2.38 As mentioned in paragraph 2.3 above, one of the investments that *ought* to be adequately recognised and rewarded is that of skill, time and effort in data governance to ensure quality data. After all, it is good quality data that will yield accurate insights during data analysis or machine learning model training. However, effort invested in, for example, maintaining accuracy and updating the information in a database – which is frequently as substantial an investment as the original collection of the dataset – will often not meet the originality threshold that copyright law requires.

2.39 By implementing the DD and the SGDR regime under it, the European Commission had aimed to (1) harmonise the rules on database protection, (2) stimulate investment in databases within the EU, and (3) safeguard the balance between the interests of database producers and users.[46] Notwithstanding the noble intention, SGDR has had a somewhat tortured treatment by the Courts of Justice of the European Union ('**CJEU**'), which has consistently interpreted the scope of the SGDR narrowly:

- In *Fixtures Marketing v Organismos prognostikon agonon podosfairou*,[47] the CJEU narrowly interpreted the Directive so that SGDR protection was limited to 'primary' producers of databases (i.e. entities whose main activity is the creation of databases). If database production was only secondary to the main activity of an entity (for example, the creation of the underlying data), it would be excluded from the protection of the SGDR.[48]
- That limitation to 'primary' producers of databases reduces the relevance and availability of the DD in a data-driven world. In a digital economy, data is the by-product of digitalisation, but it is the harnessing, curation and exploitation of data that produces insights and enables data-driven technologies like

45 Walter Arthur Copinger [*et al*], *Copinger and Skone James on Copyright* (17th ed) (London: Sweet & Maxwell, 2016) at [18-09].

46 European Commission, *Commission Staff Working Document: Evaluation of Directive 96/9/EC on the Legal Protection of Databases* (SWD(2018)146/F1), EUR-Lex (26 April 2018), at 5 (*'**Evaluation of Directive 96/9/EC**'*) <*http://edz.bib.uni-mannheim.de/edz/pdf/swd/2018/swd-2018-0146-en.pdf*> (accessed 10 June 2020).

47 Case C-444/02, [2005] ECDR 3, CJEU ('***Fixtures Marketing v OPAP***').

48 *Evaluation of Directive 96/9/EC*, above, n 46 at 3.

machine learning to power product features that were hitherto not achievable.

- In *Ryanair Ltd v PR Aviation BV*,[49] the CJEU held that the DD's mandatory rights of usage (by users of databases)[50] were "not applicable to a database which is not protected either by copyright or by the *sui generis* right",[51] and that, instead, contractual provisions could govern the use of such databases. In effect, this made it arguably preferable for database owners *not* to fulfil the requirements of *sui generis* protection under the DD, and instead to rely on contractual protections unencumbered by minimum rights for users.[52]

2.40 The SGDR model – at least as it envisioned by the DD – has also been subject to significant academic criticism.

- First, it is argued that little guidance is given regarding joint ownership of a database – be it in the case of joint authors or joint owners across different EU jurisdictions.[53] In the context of big data, these questions are of particular relevance, given that the benefits of big data are typically fully unleashed only when an entity pulls data from multiple databases that may belong to a variety of owners (human authors or otherwise).
- Second, the interpretation of "substantial investment" under Article 7(1) is still vague and difficult to apply, leading to varying standards across EU member states and, in turn, to compliance challenges for businesses operating across borders.[54] In the modern context, the ease with which databases can now be produced may mean that the *de facto* threshold for "substantial investment" has inadvertently become higher. Equally, however, it could be argued that even if a single database may be easier to produce, the quantity necessary for productive and useful capitalisation in a 'big data age' has increased.
- Third, the duration of the SGDR is potentially perpetual, since the prescribed term of 15 years may easily extend every time a new "substantial investment" is undertaken.[55] This would essentially create an exclusive property rights regime with few

49 Case C-30/14, [2014] ECLI:EU:C:2015:10, CJEU ('***Ryanair***').
50 See, for example, DD, above, n 44 Articles 6(1) and 8.
51 *Ryanair*, *id* at [39].
52 Matěj Myška and Jakub Harašta, "Less is More? Protecting Databases in the EU After Ryanair" (2016) 10 Masaryk U J L & Tech 170 at 187–188.
53 Michal Koščik and Matěj Myška, "Database Authorship and Ownership of Sui Generis Database Rights in Data-Driven Research" (2017) 31(1) Int'l Rev L, Computers & Technology 43 at 60.
54 *Evaluation of Directive 96/9/EC*, above, n 46 at [5.3.3.4].
55 Chan, "Distributing the Economic Benefits of Databases", above, n 12 at [31].

public policy limitations,[56] and represent an overcompensation for the deficiencies of the copyright regime.[57]

- Fourth, property rights protection under the SGDR model enhances existing concerns that run against the protection of databases. Such concerns typically include the monopolisation of data to the detriment of social welfare, and the merits of open-access for innovation.[58] The SGDR regime is proprietary in nature, and its introduction would be declaratory of tilting the balance in favour of the generators of databases (and against third-party access for the full utilisation of works).

2.41 The SGDR, and the two-tier system under the DD, have nevertheless been praised for protecting *all* stages in the process of creating compilations and databases. In other words, the adoption of a SGDR model would duly recognise the importance of, and the need to protect, substantial investments in the *preparatory efforts* for databases[59] (albeit still subject to the ambiguous meaning of "substantial investment"). It is also questionable whether concerns around monopolies forming would in fact come to pass, given that such arguments assume both that present legal schemes (for example competition laws) are insufficient to protect against such monopolisation, and that any particular data would be tied exclusively to a single database (i.e. that there would be no other means for a third party to access that, or equivalent, data).[60]

2.42 The European Commission has itself evaluated the effectiveness of the SGDR regime under the DD, first in 2005[61] and again in 2018.[62] Those evaluations concluded that, over time, there had been considerable harmonisation in database protections across EU member states (a key objective of the DD),[63] and that the SGDR had successfully struck an

56 Lim, "Re-defining the Rights and Responsibilities of Database Owners under Competition Law", above, n 10 at [33].

57 Chan, "Distributing the Economic Benefits of Databases", above, n 12 at [32]. See also J. H. Reichman and Pamela Samuelson, "Intellectual Property Rights in Data?" (1997) 50 Vand L Rev 51 at 137.

58 P K Yong, "Database Protection: The International Debate: Balancing Users' Rights and the Protection of Databases" [2007] 6 Mal LJ xxvii at xxxv–xxxvi.

59 David Tan, "Copyright in Compilations: Embarking on a Renewed Quest for the Human Author and the Creative Spark" (2013) 18 Media & Arts L Rev 151 at 155.

60 Yong, "Database Protection", above, n 58 at xxxiv.

61 European Commission, *DG Internal Market and Services Working Paper: First Evaluation of Directive 96/9/EC on the Legal Protection of Databases*, European Commission website (12 December 2005) <*https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57136*> (accessed 10 June 2020) ('***First Evaluation of Directive 96/9/EC***').

62 *Evaluation of Directive 96/9/EC*, above, n 46.

63 *Id* at 46.

appropriately moderate balance between the needs of database makers and users.[64]

2.43 Significantly, however, those benefits were qualitative rather than quantitative: neither evaluation found significant evidence that the SGDR had materially stimulated production of databases or investment in the European database industry.[65] Ultimately, the Commission concluded that it would be disproportionate to engage in minor reform, and that any more substantial policy intervention on the SGDR model would have to be substantiated with a strong case built upon broad consultation of stakeholders.[66]

2.44 In Singapore's context, it is unclear how far the principal benefit of the SGDR identified by the Commission – its harmonising effect – would be useful or even relevant; there is little need for harmonisation of Singapore's database protection with that of other countries (in the Southeast Asian region or beyond). Nonetheless, the DD is instructive, and its ability to protect all stages of database production is an accommodating approach worth considering – not least because, as discussed above, the current creativity approach to copyright protection cannot cover preparatory efforts.[67]

(b) United States' reluctance to adopt the SGDR

2.45 The United States' legal approach to databases has presented itself as diametrically opposed to the EU's SGDR model. The US Congress has considered various types of database protection legislation, and yet none have been fully successful in their aims – the most recent attempt now being more than a decade old.[68]

2.46 Instead, to protect their databases in the US, database owners would have to rely on a combination of various laws, including the tort of unfair competition, the misappropriation doctrine and the Computer Fraud and Abuse Act.[69] This position stems in large part from the US Supreme Court's decision in *Feist Publications, Inc v Rural Telephone Service Co*.[70] The Supreme Court clarified that the US Constitution's intellectual property clause necessitated a "minimal level of creativity" to be awarded

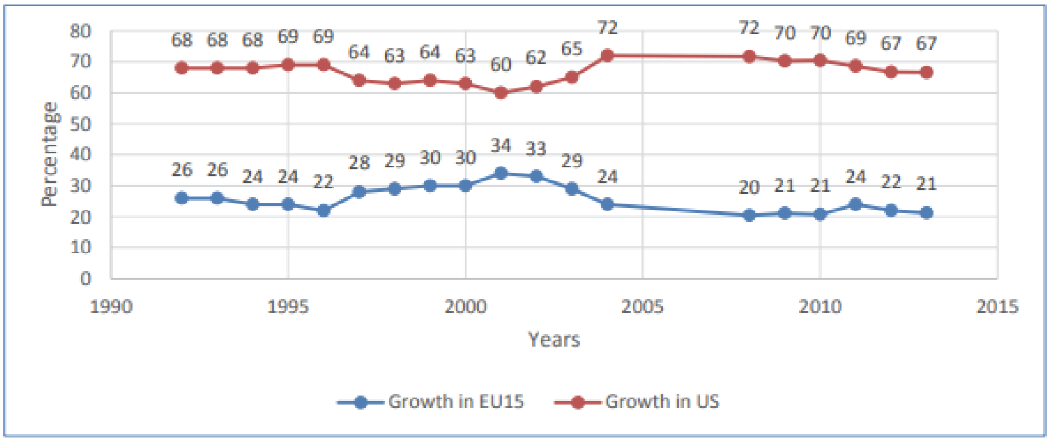
---

64 *Id* at 23.
65 *Id* at 46. See also *First Evaluation of Directive 96/9/EC*, above, n 61 at 20.
66 *Evaluation of Directive 96/9/EC*, above, n 46 at 47.
67 Tan, "New Law for Compilations and Databases in Singapore?", above, n 12 at [131], [133] and [134].
68 Marshall Leaffer, "Database Protection in the United States is Alive and Well: Comments on Davison" (2016) 57(4) Case Western Reserve L Rev 855 at 857; David F. Tamaroff, "Bottling the Free Flow of Information: A Comparative Analysis of U.S. and EU Database Protection" (2011) 12 Wake Forest J. Bus. & Intell. Prop. L. [iii] at 19–20.
69 Leaffer, above, n 68 at 855–856.
70 499 US 340 (1991) ('***Feist***').

protection.[71] In that sense, a factual compilation was beyond the protection of copyright, unless it features "an original selection, coordination or arrangement"[72] of data. In so holding, the Supreme Court affirmed that originality was a constitutional requirement, and any legislative attempt to protect databases would have to be designed to withstand such judicial scrutiny.[73]

2.47 The high constitutional threshold of originality required by *Feist*, combined with the absence of a clear legislative framework, therefore means that little formal protection is currently awarded to databases in the U.S.

2.48 Interestingly, however, this has not put the American database industry in any less of a position to compete with the European industry. Although the EU approach has resulted in higher profits for individual database owners, "it has not *grown the overall industry*, in part because it chokes off the kind of beneficial tweaking and reworking that are […] so useful to innovation".[74] The European Commission's own evaluations likewise reported more favourable tidings in the American database industry, both from 1992 to 2004 and from 2008 to 2013 (see Figure 1, below). Indeed, in its 2005 evaluation, the Commission concluded that, up to that point, the SGDR "appears to have had the opposite effect" from the growth in database production in the US, and that "the assumption that more and more layers of IP protection means more innovation and growth appears not to hold up".[75]



Source: Gale Directory of Databases 30th to 36th Edition; First evaluation of Directive 96/9/EC on the legal protection of databases.

Figure 1: Database production in the EU-15 and the United States from 1992 to 2004 and from 2008 to 2013.

71 *Id* at 358–359.
72 *Id* at 360.
73 Philip J Cardinale, "Sui Generis Database Protection: Second Thoughts in the European Union and What it Means for the United States" (2007) 6 Chi-Kent J Intellectual Property 157 at 161.
74 Kal Raustiala and Christopher Sprigman, *The Knockoff Economy: How Imitation Sparks Innovation* (Oxford: Oxford University Press, 2012) at 165–166 (emphasis in original).
75 *First Evaluation of Directive 96/9/EC*, above, n 61 at [5.2].

3 Other potential models to balance the interests of database owners and users

2.49 While ownership of databases is intended to protect the interests of their creators, it is a trite principle that the rights of ownership are not absolute. The law regularly strikes a balance between the interests of creators and the benefits of enabling access for legitimate uses. For example, in recognition of the potential for data mining to support innovation in the digital economy, the Government has proposed introducing a data mining exception to Singapore's copyright law, which would permit copying for the purpose of commercial or non-commercial data analysis, provided that the data miner gained access lawfully to the database.[76]

2.50 Apart from intellectual property or related rights associated with ownership, there are other potential legal or regulatory regimes that moderate the interests of database owners and users. This section examines various models that complement intellectual property rights in order to promote the creation and effective use of databases.

(a) Competition law

2.51 It has been suggested in some academic writings that access to databases may be further regulated under competition law. Section 47 of the Singapore Competition Act[77] prohibits unilateral conduct by a dominant undertaking amounting to an abuse of market power in Singapore – regardless of where the undertaking is geographically located.

2.52 The Competition and Consumer Commission of Singapore ('**CCCS**') has indicated that the accumulation of a large dataset is not in and of itself indicative of a firm being 'dominant' for the purpose of section 47,[78] and that the default position is that a company is not required to grant access to its data or datasets.[79] In certain defined circumstances, however, database owners that are found to be dominant in a market may contravene competition law if they anti-competitively refuse to license intellectual property rights over its databases.[80] Those circumstances are limited, but may arise if, for example, the refusal: a) relates to a so-called 'essential facility' (i.e. something for which there are no potential substitutes and

---

76 See *Singapore Copyright Review Report* (January 2019), at pp 32–34 <*https://www.mlaw.gov.sg/files/news/press-releases/2019/01/Annex%20A%20-%20Copyright%20Review%20Report%2016%20Jan%202019.pdf*> (accessed 10 June 2020).

77 Cap 50B, 2006 Rev Ed.

78 *Data: Engine for Growth*, above, n 8 at [269].

79 Nevertheless, this subcommittee recognises that there are difficulties in applying the "essential facilities" doctrine to control big data – a discussion that is beyond the scope of this report.

80 Lim, "Re-defining the Rights and Responsibilities of Database Owners under Competition Law", above, n 10 at [45].

which is indispensable to the activity in question) and b) is liable to give rise to substantial harm to competition.[81]

2.53 Competition law may thus act as a complement to copyright law in promoting the beneficial production and use of databases: while copyright protection incentivises the creation of databases, competition law prevents the assertion of intellectual property rights anti-competitively.[82] In operating *ex post*, those laws respect the protections and commercial advantage that IP rights confer on the owner, but allow for targeted intervention when denials of access cause particularly serious harm to competition in the market.[83]

2.54 In doing so, however, competition law does not have regard to the "social value" of the data at issue.[84] Concerns about personal data protection have become more relevant and acute, as there is a rising appreciation of the social implications of data on an individual's privacy, security and life. In this regard, competition law would seem unable to address the full range of considerations involved in balancing the rights of database owners, users and others.

(b) Data portability

2.55 Data portability – the legal obligation to comply with data subject requests for their personal data to be moved from one organisation to another – has been identified as a possible regulatory tool to address competition and other access-related concerns in big data.[85] Such portability has evident potential to help address costs and barriers for consumers in switching between service providers and so to reduce risks of customer 'lock-in'. But equally importantly, it also has the potential to

81 See *CCCS Guidelines on the Treatment of Intellectual Property Rights in Competition Cases 2016*, CCCS (2016) <*https://www.cccs.gov.sg/-/media/custom/ccs/files/legislation/legislation-at-a-glance/cccs-guidelines/cccs-guidelines-on-the-treatment-of-intellectual-property-rights-in-competition-cases.pdf*> (accessed 10 June 2020) ('***CCCS Guidelines***'). This 'essential facilities' doctrine is modelled on that developed under EU law.

82 Chan, "Distributing the Economic Benefits of Databases", above, n 12 at [38] and [39].

83 *Id* at [77].

84 Catherine Colston, "Challenges to Information Retrieval – a Global Solution?" (2002) 10(3) Int'l J L & Info Tech 294.

85 *Data: Engine for Growth*, above, n 8 at [240] and Annex 2, and Personal Data Protection Commission in collaboration with the Competition and Consumer Commission of Singapore, *Discussion Paper on Data Portability* (Singapore: Personal Data Protection Commission, 2019) at [3.1] *et seq* <*https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/Data-Portability/PDPC-CCCS-Data-Portability-Discussion-Paper—250219.pdf*> (accessed 10 June 2020) ('**Discussion Paper on Data Portability**'); and *Public Consultation on Review of the Personal Data Protection Act 2012 – Proposed Data Portability and Data Innovation Provisions* (Singapore: Personal Data Protection Commission, 2019) <*https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Legislation-and-Guidelines/PDPC-Public-Consultation-Paper-on-Data-Portability-and-Data-Innovation-Provisions-(220519).pdf*> (accessed 10 June 2020).

enable the producers of new services – whether established businesses or start-ups – to have access to data. It could therefore help drive external benefits from increased data use, higher productivity and recombinant innovation.[86]

2.56 To this end, the Personal Data Protection Commission ('**PDPC**') intends to introduce a data portability obligation within the Personal Data Protection Act ('**PDPA**')[87], having publicly consulted on this proposal.[88] In short, the obligation – which would cover any organisations that operate in Singapore – applies to "user provided and user activity data of individuals with whom the porting organisation has a direct and existing relationship",[89] but not to any "derived data".[90] Organisations would not be required to provide a copy of that data to the data subject themselves,[91] or to an entity with no presence in Singapore.

(c) Tort of unfair competition

2.57 This doctrine arose from the US Supreme Court case of *International News Services v Associated Press*.[92] Under this cause of action, a party may hold a "quasi-property" interest in information as against a competitor, so long as that piece of information contained economic value.[93] This may in turn be applied *ex post* for the protection of databases: the fairness of the reuse of data would have to be judged by the courts, taking into account factors such as the amount of data appropriated, the nature of such data and the purpose for the appropriation.[94]

2.58 The distinct merit of such a model would be to allow the judiciary to modify the level of intervention and protection on a case by case basis, in accordance with the social value of the database.[95] However, despite the benefits of such individualised justice and welfare maximisation, in practice

86 *Discussion Paper on Data Portability*, above, n 85 at [3.1] *et seq.*
87 No 26 of 2012.
88 See Personal Data Protection Commission, *Response to Feedback on The Public Consultation on Proposed Data Portability and Data Innovation Provisions* (Singapore: Personal Data Protection Commission, 2020) <*https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Legislation-and-Guidelines/Response-to-Feedback-for-3rd-Public-Consultation-on-Data-Portability-Innovation-200120.pdf*> (accessed 10 June 2020) ('**Data Portability Consultation**'); and Public Consultation on the Draft Personal Data Protection (Amendment) Bill <*https://www.mci.gov.sg/public-consultations/public-consultation-items/public-consultation-on-the-draft-personal-data-protection-amendment-bill*> (accessed 10 June 2020).
89 *Data Portability Consultation*, above, n 88 at [3.8].
90 'Derived data' being any "new data element that is created through the processing of other data by applying business-specific rules" *Id.* at [3.3].
91 This is covered under the extant data subject access obligation under section 21 of the PDPA.
92 248 US 215 (1918).
93 *Id* at 234.
94 Chan, "Distributing the Economic Benefits of Databases", above, n 12 at [45].
95 *Id* at [41] and [46].

the process of having to judicially vindicate a database is time and resource consuming.[96] Each case would require a detailed and multi-factorial consideration, the time-cost and expense of which may well outweigh the benefit it proffers – especially in a business landscape as fast-paced and dynamic as that in which database and big data innovation typically occurs.

(d) Contract

2.59 The commercial reality today is that parties typically resort to contractual arrangements to ensure adequate protection of databases. Significantly, however, the focus of such contractual protection is the contracting parties – it does not protect against third-party infringement. For this reason, the EU's 2018 evaluation of the SGDR noted that contract laws are complements, rather than alternatives, to the SGDR model.[97]

2.60 Contract law also has its own limitations. First, contracts governing the use of databases are often standard form contracts (usually in favour of the database owner) rather than freely negotiated ones.[98] Second, and conversely, even if a contract is established with users of the database, the digital environment makes it near-impossible for database owners to monitor their use and enforce contractual provisions.[99] The ability of contract law to manage the rights and obligations of database owners and their users thus assumes, in effect, the existence of intellectual property rights in the database.

C RECOMMENDATIONS

1 No real impetus for Singapore to introduce a sui generis right

2.61 From a purely analytical perspective, it may be intellectually appealing for Singapore to consider conferring a two-tier database protection system in a similar fashion to the EU's DD – namely (1) copyright to protect the structure of the database, and (2) a *sui generis* right to protect databases – particularly in light of the Court of Appeal's clear statement in *Global Yellow Pages* as to the limited protection offered by existing copyright laws.

2.62 After the initial design and implementation of a database, there is doubt at present as to whether copyright extends to the subsequent

96 *Id* at [47].

97 *Evaluation of Directive 96/9/EC*, above, n 46 at 31–32, but note the earlier observation of the counter-intuitive effect of *Ryanair* on weakening the SGDR model.

98 Although where the database user is contracting as a consumer, the Unfair Contract Terms Act (Cap 396, 1994 Rev Ed) may offer additional protections against the excesses of database owners, such as unfair exclusion of liability or unfair indemnity clauses.

99 Yong, "Database Protection", above, n 58 at lxx.

routine efforts in data entry or collection, or to other activities for ensuring data quality, all of which require significant investments and labour. The concern is that organisations may not be incentivised to invest efforts into building databases if they are not sufficiently rewarded for it.

2.63 However, from an economic perspective, the EU's SGDR does not appear to have grown the database industry in Europe. That absence of a clear economic impact, coupled with the uncertainties in defining an adequate legal threshold ("substantial investment" in the EU context) and framework (for example, joint ownership and authorship), point to the challenges of designing such a right and militate against its adoption in Singapore.

2.64 Indeed, even more fundamentally:

(a) It is questionable whether there is presently any issue in Singapore with organisations not having the incentive to invest in databases.[100]

(b) Arguments have been made that, in a big data age, not extending the current monopoly rights of database creators would better promote innovation overall, given the complexity of, and the interaction of players in, the data landscape.[101]

2.65 In view of the foregoing, we are of the view that there may not be a real impetus for Singapore to introduce a *sui generis* database right. Additionally, as Singapore seizes the opportunities to be a Smart Nation, we have concerns as to whether introducing of such a right would encourage or impede potential innovations in data sharing.

2.66 While this may leave a lacuna (as revealed in *Global Yellow Pages*), the EU experience shows that there may be little economic benefit (e.g. in the form of increased investment in databases) in seeking to fill this by introducing a standalone right. Organisations can continue to rely on the existing range of legal rights available to complement copyright protection of their electronic databases, including contract (for example, terms of use to restrict copying of data),[102] trade secrets and confidential information.

2.67 Instead, we consider that it may be useful to clarify – either through IPOS administrative guidance (or similar 'soft law' measures) or subsidiary

---

100 Trina Ha and Gavin Foo, "Monopoly Rights vs Freedom of Access: The Copyright Balance in a Data-driven Economy" [2018] Personal Data Protection Digest 57 at [28]; and Wolfgang Kerber, "A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis" (2016) 11 Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil [GRUR Int] 989.

101 Ha and Foo, *id* at [29].

102 Lau Kok Keng, Nicholas Lauw and Jiamin Leow, *The Challenges of Protecting a Database without a Sui Generis Right, this Time from Singapore*, The IPKat (1 August 2017) <*http://ipkitten.blogspot.com/2017/08/the-challenge-of-protecting-database.html*> (accessed 10 June 2020).

legislation – *how* and *when* electronic databases can enjoy compilation rights under Singapore's copyright regime. As explained above (paragraphs 2.15 to 2.27) there are real practical difficulties associated with applying the existing compilation rights analysis to electronic databases.

2.68 We believe that greater clarity and more explicit recognition of the significant intellectual effort required to implement efficient and scalable electronic databases, would be welcome in at least the following areas:

(a) Clarifying how copyright principles like the idea/expression dichotomy and the merger doctrine[103] apply to electronic databases (for example, schemas and tables).[104]

- Database design – no differently from software design – is guided by design principles, technical limitations and the need to ensure efficiency in utilisation of computing resources. Within these boundaries, there can still be room for expressing similar ideas differently in ways that meet the threshold for copyright protection. And while it might be argued that short lines of codes that perform common or simple database queries should be treated as utilitarian under the merger doctrine, those lines of code form units within the electronic database which, as a whole, would usually still be sufficiently expressive.

(b) Clarifying how the element of *selection* is met by particular activities, such as a software designer translating business rules (for example, what data, where do you obtain the data) into software code.

- Business rules are translated into software code or database rules by software designers (e.g. identifying the parameters before a reading from a sensor is selected for storage, imposing rules to reject data from certain ranges, etc). These should amount to sufficient originality for selection, notwithstanding that they are pre-determined and coded into the application system before it is deployed and data is selectively recorded.

(c) Clarifying how the element of *arrangement* is met by, for example, the software designer's decisions in creating the

103 That is, the principle that "expression is not protected … where there is only one or so few ways of expressing an idea that protection of the expression would effectively accord protection to the idea itself" (*Bellsouth Advertising & Publishing Corporation v Donnelley Information Publishing*, Inc 999 F 2d 1436 (1991) (US Court of Appeals, 11th Circuit) at 1442).

104 Daniel Gervais, "Exploring the Interfaces between Big Data and Intellectual Property Law" (2019) 10(1) J of Intellectual Property, Information Technology & E-commerce Law at [28] <*https://www.jipitec.eu/issues/jipitec-10-1-2019/4875/JIPITEC_10_1_2019_3_Gervais_Big_Data_IP*> (accessed 10 June 2020)

relationships between database tables in a relational database.

- These relationships are crucial to good database design in order to ensure scalability and reduce latency in database queries. The fact that they are not visible to the unaided human eye does not detract from their nature as works that can be original and expressive, tailored as they are to ensure that the application system that the electronic database supports can operate effectively and efficiently.

(d) Clarifying how, when an electronic database is in use, copyright protects the selection of records based on user input and their presentation to the user through the graphical user interface of the application system.

- While the selection of records for display and their arrangement on screen is purely for the purpose of presentation – and not for creation of the underlying compilation right – there is the potential that the selection and arrangement criteria are sufficiently expressive that a derivative work is extracted, i.e. a separate compilation.

(e) Clarifying how 'sweat of the brow' efforts in data entry or collection and the routine and laborious activities involved in ensuring data quality in the database can meet the originality requirement for compilation rights.

- This requires getting into the details of database maintenance after its initial setup and during its life span. Proper data management will entail the application of business rules towards data-entry, as well as database maintenance activities to clean up records within the electronic database to ensure sufficient accuracy, completeness and currency for their intended use. These are activities that take place regularly and which the EU SGDR sought to protect separately from compilation rights.
- While there is no strong case for the creation of a new statutory regime just for this purpose, it would nevertheless be useful to clarify how Singapore's common law 'sweat of the brow' approach to copyright protection will protect this sort of investment into data-entry and database maintenance.
- Another related area for guidance is how records of such activities, which are typically undertaken by several different members of staff, ought to be kept in order to meet the evidential requirement of proving human authorship.

2.69 This report acknowledges that the proposed clarifications would not expand protection conferred on owners of electronic databases, but merely clarify the applicability of copyright protection for electronic databases. However, we consider that with these clarifications, whether made through IPOS administrative guidance (or similar 'soft law' measure) or subsidiary legislation, the gaps that the SGDR was intended to address would also be effectively addressed within Singapore's copyright regime. Furthermore, how compilation rights are created when electronic databases are designed, deployed and operated could also have been clarified.

2.70 After these interventions, it is our view (taking into account the competing societal goals referred to at paragraph 2.3 above) that the existing rights conferred on electronic database owners (who are usually able to avail themselves of larger resources) are adequate and need not be expanded. We consider that this approach will adequately protect database owners from both unauthorised copying and the wrongful appropriation of expression for financial gain,[105] and is recommended after careful consideration of the countervailing need to encourage a vibrant data economy.[106]

2 Re-examine the fundamentals of authorship under Singapore's Copyright Act

2.71 Another emerging trend is the deployment by organisations of complex machine learning algorithms that are capable of generating their own databases. Amazon's Deep Scalable Sparse Tensor Network Engine,[107] for example, tailors Amazon users' experiences with unique, personalised displays and targeted recommendations based on the *machine's* (rather than a human's) recognition of users' preferences and purchase histories. Alternatively, even where humans do design and develop the electronic databases, modern technologies mean that data collection could be fully automated, for example, through IoT sensors.

2.72 This report acknowledges that even with the proposed clarifications suggested above, it is questionable whether organisations such as Amazon would enjoy protection in the output (namely, the specific webpage listings for each user) of a complex machine learning algorithm. Similar ambiguities arise for databases that are compiled through automated means. As such,

105 See *I-Admin (Singapore) Pte Ltd v Hong Ying Ting and others* [2020] SGCA 32.

106 See Kenneth Cheng, "The Big Read: Strides Made, but Some Way to Go for Government to Quench Thirst for Data", *Today* (13 July 2019) <*https://www.todayonline.com/big-read/big-read-strides-made-some-way-go-government-quench-thirst-data*> (accessed 10 June 2020), where it was highlighted that start-ups' access to public data by the Singapore Government has resulted in significant innovations.

107 Kiuk Chung, "Generating Recommendations at Amazon Scale with Apache Spark and Amazon DSSTNE", AWS Big Data Blog (9 July 2016) <*https://aws.amazon.com/blogs/big-data/generating-recommendations-at-amazon-scale-with-apache-spark-and-amazon-dsstne/*> (accessed 10 June 2020).

we consider that it may be useful for Singapore to revisit the fundamental requirement for a human author for a copyright work.

2.73 We note that, in the United Kingdom, the Copyright, Designs and Patents Act 1988 (‘**CDPA**’)[108] provides for the notion of “computer generated” works, which have no human author.[109] For literary, dramatic, musical or artistic works that are computer generated, the author, for copyright purposes, is taken to be “the person by whom the arrangements necessary for the creation of the work are undertaken.”[110] Equivalent provisions also exist in copyright legislation in India,[111] New Zealand[112] and Hong Kong.[113] However, in the most recent Singapore copyright review conducted from 2016 to 2019,[114] this issue was not raised for deliberation.

2.74 By acknowledging that copyright should subsist even when there is no human author of a literary, dramatic, musical or artistic work, the UK CDPA both:

(a) offers potential copyright protection to organisations such as Amazon that use complex machine-learning algorithms to computer-generate literary works (i.e. databases), and

(b) provides greater rules for determining authorship where data collection is fully automated.

2.75 It is true that there are ongoing academic debates in the UK as to whether the CDPA provisions create ambiguity as to the identity of the author, given that authorship of the work could be attributed to the programmer, end user, machine-generated algorithm (as a *non persona*), or to no one at all. However, we consider this issue can be resolved on a fact-specific analysis of who directed the machine-generated algorithm to produce the work in question.[115]

2.76 With further advances in AI technologies, e.g. active learning machine learning models, the day will soon arrive when the human input into a work may be seen as so remote that human authorship is lost. As such we believe it may be useful for Singapore to further consider reforms in this area by adopting section 9(3) of the CDPA. At the very least, we see

108 1988 c 48 (UK).
109 CDPA, s 178.
110 CDPA, s 9(3).
111 Copyright Act 1957 (No 14 of 1957; India), s 2(d)(vi).
112 Copyright Act 1994 (1994 No 143; NZ), s 5(2).
113 Copyright Ordinance (Cap 528; HK), s 11(3).
114 Ministry of Law and Intellectual Property Office of Singapore, *Singapore Copyright Review Report*, Ministry of Law website (17 January 2019) <*https://www.mlaw.gov.sg/files/news/press-releases/2019/01/Annex%20A%20-%20Copyright%20Review%20Report%2016%20Jan%202019.pdf*> (accessed 10 June 2020).
115 Andrés Guadamuz, “Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works” (2017) 2 Intellectual Prop Qtly 169.

benefit in issuing (by way of subsidiary legislation or IPOS administrative guidelines) guidance as to the circumstances, if any, where computer-generated works can enjoy copyright protection.

3 Concluding observations

2.77 Assessment of the suitability of any alternative database protection models must take into account the particular features of Singapore's context:

(a) First, the absence of any international consensus on database protection means there is limited scope for Singapore to harmonise its laws with others. Indeed, it may wish to design a database protection model specifically to make it attractive for database creators to base their operations and storage in the jurisdiction.

(b) Second, cognisance should nonetheless be had to existing legal frameworks and doctrines when adopting any given model. Even if coherence is not an overriding policy consideration, attempts to introduce a new regime (such as the SGDR) should be done cautiously and reservedly.

2.78 The nature of data use and exploitation is changing. Big data involves the use of technology and analytical algorithms to derive value, and goes beyond the mere compilation and arrangement of individual records of data. Regardless of the approach taken (i.e. creation of new protections or adjustment of existing rules), it is vital that legal frameworks applicable to databases which engage in big data-type analysis recognise this distinction.

CHAPTER 3

DATA OWNERSHIP

3.1 Closely connected to the issue of database protection is the status of rights granted over data (as opposed to databases). In the face of significant collection of data (whether voluntary or not), in particular from IoT devices, it is appropriate to consider whether property rights (as opposed to privacy obligations) ought to be accorded to data, given the sensitivity of, and the value that can be extracted from, such data.

3.2 At the outset, it is appropriate briefly to reiterate that when we speak here of ‘data’, we are referring to either an individual data element or a combination of several data elements that forms a record describing a single instance. In the context of datasets or databases, that ‘data’ would constitute a single record in the dataset or database, e.g. one person’s entry in a nominal roll or a single entry recording the temperature at a particular date and time taken from a sensor in a meeting room.

3.3 Presently, there is some acknowledgement that data processors and/or intermediaries have legal control over databases. What is unclear is whether an individual (i.e. the data subject) has sufficient control and ‘ownership’ over the individual records that make up the databases. If property rights do not subsist in the data, then – unlike, for example, with land – there can be no “rights of possession, use and enjoyment, which the owner can bestow, collateralise, encumber, mortgage, sell or transfer, and the right to exclude everyone else from”.[116] Consequently, the ‘owner’ can neither have his or her data stolen (in a property sense),[117] nor enjoy a right of access to the medium on which the data is stored.[118]

3.4 The present regime in Singapore contemplates that if the data involves identifiable information about an individual (termed ‘personal data’, and discussed further below), then it will enjoy certain protection under the PDPA. The PDPA confers on a specific data subject rights over his personal data.[119] It does not, however, confer legal ownership of the data, in the manner that, for example, the intellectual property rights regime confers legal ownership of patents, copyrights and trademarks.

---

116 Directorate-General of Communications Networks, Content and Technology, European Commission, *Legal Study on Ownership and Access to Data: Final Report* (Luxembourg: Publications Office of the European Union, 2016) at 6 (‘**Legal Study on Ownership and Access to Data**’) <*https://publications.europa.eu/en/publication-detail/-/publication/d0bec895-b603-11e6-9e3c-01aa75ed71a1*> (accessed 10 June 2020).

117 As seen in *Oxford v Moss* (1979) 68 Cr App Rep 183, Div Ct (England & Wales).

118 As seen in *Attheraces Ltd v British Horseracing Board* [2005] EWHC 1553 (Ch), HC (England & Wales).

119 For example, the rights to access and correction (ss 21 and 22 of the PDPA) and the future right to data portability: see above, at [2.56].

3.5 The lack of clarity over legal ownership of discrete records (as opposed to databases) may lead to mistaken assumptions in the marketplace. While businesses in Singapore may seek to resolve ownership issues surrounding data transferred to them through express provision in contracts, there is a risk that they may perceive data as a property right, and assume they can rely on implied property law concepts (such as the transferor being implied to have given certain covenants for title) when such assumptions may be suspect.[120]

3.6 In this part, therefore, we consider how ownership rights might be applied in the context of data, and whether it would be practicable or desirable for Singapore law to ascribe such rights, or rather to continue to rely on concepts around who controls the data, who has custody of it and individuals' ability to assert rights over how their personal data is used and shared.[121]

A CLASSIFICATIONS OF DATA

3.7 There is no universally accepted taxonomy of data. Rather, data can be classified in a variety of manners: human versus machine-generated; quantitative versus qualitative; discrete versus dynamic, and so on. However, for present purposes, we consider it appropriate to distinguish between personal data and non-personal data, in particular given the importance of privacy in today's climate.

1 Non-personal data

3.8 Non-personal data refers to all data that do not identify an individual. These include a whole range of data, including financial market data, operational data, and market research data. In the digital economy of IoT devices and AI tools, such non-personal data has numerous commercial applications, and plays a significant role in, for example, allowing organisations to understand trends and analysis.

120 *Legal Study on Ownership and Access to Data*, above, n 116 at 7.

121 In its recent report on AI, a UK House of Lords Select Committee rejected the data ownership approach in favour of a data control approach: House of Lords Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?* (HL Paper 100), Parliament of the United Kingdom website (16 April 2018) at [62] ('**Lords Committee AI Report**') <*https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf*> (accessed 10 June 2020). See in particular, Select Committee on Artificial Intelligence, *Corrected oral evidence: Artificial Intelligence - Tuesday 31 October 2017* at [Q56] <*http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/oral/73546.html*> (accessed 10 June 2020).

3.9 The key stakeholders who will likely value such non-personal data are governments and organisations.[122] Such stakeholders would have invested significant resources in obtaining non-personal data, and are able to rely on both legal and technical measures to ensure that they maintain control of that data:

- The majority of non-personal data are facts, e.g. readings obtained from IoT sensors, statistics collated from surveys, spot prices or indices derived from market activities, etc. Copyright law has little, if any relevance in this context, as there is typically no 'expression' for it to protect or – even if there is – the merger of expression and facts renders any originality nugatory. However, in Singapore, such non-personal data could still be protected by the law of confidence, or through a range of sectoral legislations (for example, the Official Secrets Act[123] with respect to state information).
- Governments and organisations can also implement technical and security measures such as preventing users from easily porting data to another system, watermarking of data, and/or audit trails.[124]

3.10 The value of non-personal data is not in the individual records of facts but in the compilation of many of these records in a database. Ambient temperature records have little value individually, for example, but a dataset of ambient temperature of many meetings rooms, over an extended period of time, could be extremely valuable to both:

(a) the premises owner (in identifying when the rooms are typically not in use, how often they are used, which days and times are busiest, etc); and

(b) the manufacturer of the smart thermostat (in helping to improve a machine learning-based AI model used to predictively adjust the level of the thermostat according to whether the room is in use).

3.11 The data generated by the temperature sensor therefore has potential value, which is fully realised only in its aggregation. While they

122 Lords Committee AI Report, above, n 121 at [64]–[66], where the Select Committee noted that tremendous amounts of data are being collected by large technology companies (such as Alibaba, Alphabet, Amazon, Apple *etc*) and governments.

123 Cap 213, 2012 Rev Ed.

124 British Academy, Royal Society and techUK, *Data Ownership Rights and Controls: Reaching a Common Understanding: A Discussion at a British Academy, Royal Society and techUK Seminar on 3 October 2018*, Royal Society website (3 October 2018) at 18 ('**Data Ownership Rights and Controls**') <*https://royalsociety.org/-/media/policy/projects/data-governance/data-ownership-rights-and-controls-October-2018.pdf*> (accessed 10 June 2020).

may have different or competing interests and commercial objectives in tapping into the same data stream (of ambient temperature readings in this example), the value to both the facility owner and the manufacturer of the smart thermostat is in the dataset that is compiled and how each of them analyses it.

- From a technical perspective, it is possible for both to tap into the data stream although there are practical difficulties. For example, the smart thermostat manufacturer may not have built Application Programming Interfaces (APIs) that allows another software to read the temperature sensor in the smart thermostat, or there may be restrictive contractual clauses that prevent such activities, even when the facilities owner has the technical knowhow and ability to do so.
- Apart from technical challenges, there may be practical commercial challenges arising from an imbalance in bargaining power between the parties, e.g. the smart thermostat manufacturer may not have any incentive to deal with requests for access from facilities owners who have purchased their products.

3.12 The existing legal framework in Singapore pertaining to non-personal data generally ensures that entities that created data, or are able to control how data is used, retain sovereignty over their data and, to date, there has not been significant, widespread call for reform. Nonetheless, in the light of the above, we consider that there is something to be said for the introduction of a right not dissimilar to data portability (as discussed at paragraphs 2.55 to 2.56 above, and further below), but centred on non-personal data.

3.13 Specifically, such a right could serve to address the imbalance of bargaining power described above by creating a legal impetus to offer APIs that enable alternative uses of data or to respond to legitimate access requests:[125]

- As alluded to above, the typical tension that arises in relation to non-personal data will be between a manufacturer and its customer (usually a business entity and not an individual in his personal or domestic capacity).
- The data is likely to be generated through usage of devices or operation of equipment and machinery, and thus the question of data accuracy is less of an issue (or more likely to be

125 It should be borne in mind that, insofar as access to the dataset that has been compiled by the smart thermostat manufacturer from the data stream is concerned, the earlier discussion on competition law is pertinent: see above, paragraphs 2.51–2.54.

technical in nature, e.g. related to calibration and measurement bias).

- Without the customer, there would be no usage data. However, at present, the manufacturer has the ability to control access – and thus lock in customers – while utilising the data to improve the efficiency and accuracy of its product.
- Clarification that the customer also has the right to gain access to and acquire a copy of the data generated by its use of the product could therefore serve similar policy objectives to the right of access and portability that have been granted to data subjects over their personal data.

3.14 This is the direction that has been taken in Australia, for example, with the introduction of a consumer data right that applies to both individuals and businesses.[126] We note also that, for its part, the EU has implemented Regulation (EU) 2018/1807,[127] a framework for the free-flow of non-personal data in the EU. This Regulation further strengthens the sovereignty of a person or entity over their non-personal data, by a) prohibiting Member States from implementing data localisation requirements (unless due to public security reasons), and b) taking steps (for example, data portability obligations) to prevent users of data processing services being locked-in to those services by private contractual, legal and technical restrictions.

2 Personal data

3.15 Personal data presents an added dimension compared to non-personal data: the individuals' identifying information is at stake. The interests of the individual data subjects are often not aligned to those of organisations; while individuals may desire – and instinctively expect – 'ownership' and control over their personal data that they may disclose to organisations, the balance of power and resources often lies in favour of those organisations.[128]

126 The Australian consumer data right is currently under pilot in the banking sector. See the Australian Competition & Consumer Commission's web pages on Consumer Data Right: <*https://www.accc.gov.au/focus-areas/consumer-data-right-cdr-0*> (accessed 10 June 2020).

127 Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a Framework for the Free Flow of Non-personal Data in the European Union, OJ L 303 28.11.18, p. 59 <*https://eur-lex.europa.eu/eli/reg/2018/1807/oj*> (accessed 10 June 2020).

128 *Data Ownership Rights and Controls*, above, n 124 at 18.

3.16 Against that backdrop, calls for granting property rights over personal data are beginning to gain prominence in popular media,[129] especially in the wake of scandals such as the wilful exploitation of Facebook data by Cambridge Analytica. Given the extent to which personal data is increasingly used by organisations to generate profits, there is for many an intuitive attraction to granting property rights over personal data, so as to preserve a degree of individual privacy and agency in today's interconnected reality. Certainly, the merits of such a right at least warrant consideration.

B MERITS OF GRANTING PROPERTY RIGHTS OVER *PERSONAL* DATA

3.17 There are various arguments that can be advanced to support the creation of a property rights regime for data.

3.18 First, such a regime would, at first sight, appear to offer a clear and coherent means to protect privacy. Our individual instinct to control 'our' personal data, which we often regard as an extension or even part of ourselves, leads us to reach out to ownership rights in property as a framework to assert control. This is largely because of the ability for property rights to enforce control (over personal data) against the entire world (otherwise known as the *erga omnes* effect).[130] Indeed, commentators have taken the view that property provides an invaluable "set of rules that enables legal owners to share the benefits of their assets with third parties by way of derivative interests".[131] A regime for data based on property rights would be based on well-established principles and the regime could be developed by drawing analogous concepts from the law of property.

3.19 Second, such a regime may align with the European Convention of Human Rights ('**ECHR**') jurisprudence on the *public law* notion of 'property'. The European Court of Human Rights ('**ECtHR**') has developed a distinct notion of 'possessions' for the purposes of Article 1 of the First Protocol of the ECHR, which gives the notion of 'property' a wider ambit under public law than that found under private law.[132] The crucial factor for determining whether an intangible entitlement is a 'possession' under

129 Brittany Kaiser, "Facebook should pay its 2bn users for their personal data", *Financial Times* (9 April 2018) <*https://www.ft.com/content/7a99cb46-3b0f-11e8-bcc8-cebcb81f1f90*> (accessed 10 June 2020).

130 Nadezhda Purtova, "Do Property Rights in Personal Data Make Sense after the Big Data Turn? Individual Control and Transparency", Tilburg Law School Research Paper No 2017/21 at 6. <*https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3070228*> (accessed 10 June 2020).

131 *Data Ownership Rights and Controls*, above, n 124 at 13.

132 Tom Allen, *Property and the Human Rights Act 1998* (Oxford: Hart Publishing, 2005), at 40-41.

Article 1 is the existence of some economic value and its marketability.[133] Data, as a species, arguably satisfies both these criteria.

3.20 Lastly, there is an increasing realisation that individuals' personal data are being enclosed and harnessed by a few technology giants.[134] A property rights regime may offer a defensive measure to hedge against the exploitation-by-enclosure of personal data for profit by a few large corporations, and limit incursions on the privacy and agency of individuals.[135] By acknowledging a different starting point – that data subjects own their data – such a regime may allow for a better allocation of the value in such data.

1 The existing legal framework

3.21 Given these potential benefits, it is appropriate to consider whether property rights – or (given that substance is more important than form) rights substantively akin to property rights – can be found in existing parts of the Singapore legal framework. Specifically, in laws that have, in one form or another, information (or data) as their subject matter – for example, copyright, confidentiality, privacy and data protection law, each of which is considered below.

(a) Copyright, confidentiality and privacy

3.22 As previously noted in relation to non-personal data, it is clear that *copyright law* accords little protection over individual strands of data or information. The law of copyright seeks to protect only the 'expression of ideas' and not the idea *per se*.[136] This is reflected in the requirement that an idea has to be fixed in a material form for it to be recognised as a 'work' worthy of copyright protection.[137]

3.23 Similarly, the *law of confidentiality* (which is rooted in the equitable doctrine of breach of confidence) affords little viability as an alternative to protect such individualised data or information, or to confer rights equivalent to property rights. Specifically, it is "widely accepted" that protection of confidential information afforded by the doctrine of breach of confidence rests on the relationship between parties, and is irrespective of

133 Tanya Aplin, "Confidential Information as Property?" (2013) 24(2) King's LJ 172 at 178.
134 Purtova, above, n 130 at 10.
135 *Id* at 11.
136 Susanna H S Leong, "Law of Confidence" in *Intellectual Property Law of Singapore* (Singapore: Academy Publishing, 2013), 1133 at [39.051].
137 Copyright Act, above, n 22, s 16(1); Berne Convention for the Protection of Literary and Artistic Works (as amended on 28 September 1979; in force on 19 November 1984), Art 2(2); Leong, "Historical Origins and Current State of Copyright Laws" in *Intellectual Property Law of Singapore*, *id* at [03.081], and Leong, "Law of Confidence", *id* at [39.051]. See also Ha and Foo, above, n 100 at [13].

deliberations of property rights.[138] In Singapore, it has been noted that the law of confidentiality operates either a) to respect the confider's right of privacy, b) to enforce a contractual term that disclosure would be confidential, or c) on the basis of good faith owed between parties.[139] All those concepts notably avoid the notion of property rights.

3.24 There are even challenges to seeking to use the *law of privacy* to ascribe quasi-property rights over information, notwithstanding its particular focus on private information and its indistinct (and thus, in principle, expandable) boundaries:[140]

- Regardless of whether one favours the European understanding of privacy (focused on protecting the "honour or dignity of individuals") or the American conception (which sees privacy primarily as protecting a liberty interest.),[141] privacy law is viewed as the guardian of the individual, rather than being rooted in the information at hand, or its qualities.
- The law of privacy in Singapore (as distinct from data protection law) is in its infancy.[142] In Singapore, privacy is viewed as an exercise in trade-offs, where Singaporeans accept, for example, certain state interferences (in the form of national ID card programs and surveillance activities) in exchange for security and material comfort.[143] The drive has been rooted more in the economic imperatives of globalisation, than in human rights[144] or notions of individualism.

138 Leong, *id* at [39.030]. See also *Fraser v Evans* [1969] 1 QB 349 at 361, CA (England & Wales); *Wheatley v Bell* [1984] FSR 16, SC (NSW, Aust); and Aplin, "Confidential Information as Property?", above, n 133. In *Douglas v Hello! Ltd (No 3)* [2008] 1 AC 1 at [300], HL (UK), Lord Walker furthermore warned against over-extension of the breach of confidence doctrine: "the law of confidentiality [should not] afford the protection of exclusivity in a spectacle […] [This] would in effect confer on the exclusive licensee a form of property rights which the courts have […] rightly withheld".

139 Ng-Loy Wee Loon, *Law of Intellectual Property of Singapore* (2nd ed) (Singapore: Sweet & Maxwell Asia, 2014) at [38.1.1]–[38.1.2].

140 Lanx Goh and Jansen Aw, "Digital Protection Law and Privacy in Singapore" in Simon Chesterman (ed), *Data Protection Law in Singapore: Privacy and Sovereignty in an Interconnected World* (Singapore: Academy Publishing, 2018), ch, 4 at [4.4]. See also Gilbert Leong, Foo Maw Jiun and Kenneth Fok, "Protecting the Right of Publicity under the Personal Data Protection Act" 293 [2017] PDP Digest <*https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/pdp-digest-2017–270717.pdf*> (accessed 10 June 2020).

141 Simon Chesterman, "From Privacy to Data Protection" in *Data Protection Law in Singapore*, *id*, ch. 2 at [2.17].

142 See generally, *My Digital Lock Pte Ltd* [2018] SGPDPC 3.

143 Goh and Aw, "Digital Protection Law and Privacy in Singapore", above, n 140 at [4.10]; and Adrian Tan, *Privacy is Dead*, TSMP Law website (4 June 2018) <*http://www.tsmplaw.com/forefront/privacy-is-dead/*> (accessed 10 June 2020).

144 Chesterman, above, n 141 at [2.28]–[2.29].

- Additionally, the patchwork nature of the doctrines and causes of action that underpin the law of privacy serve to stress this point. It has been noted that "while it is probably still true that the common law does not recognise a general right to privacy, there exists today a framework of common law and statutory torts that collectively protect an individual's privacy".[145] For example, there is a developing tort of misuse of private information in the United Kingdom, first pronounced by the House of Lords in *Campbell v MGN Ltd* [2004] 2 AC 457.[146] This was pronounced without a general right to privacy being recognised by the English Parliament. Prior to such a tort, the doctrine of breach of confidence, discussed above, was stretched to cover such concerns as the notional infringement of privacy.
- Such a patchwork and piecemeal framework would appear at odds with a conception of privacy law as being capable of according quasi-property rights over information. The key aspect of a property right, notwithstanding its own amorphous definitions, is that it is good against the world (namely, a right *erga omnes*). Seeking to 'construct' such *erga omnes* protection using developing privacy laws would threaten to disrupt their growth and distort their final form. At a stage when privacy law is in its infancy in Singapore, it would be unrealistic to expect it to solve the vagaries of ownership over personal data.

(b) Data Protection and incidents of ownership

3.25 For those seeking assurance that there is a branch of law that confers control over data, personal data protection law is perhaps the best place to look, even if the rights it confers currently only adhere to personal data.

3.26 Personal data is information *about an individual* who can be identified from that information (or from its combination with other information that an organisation has access to).[147] The focus of personal data protection – in contrast to compilation or database rights – is on the individual record (although the obligations imposed by the PDPA on data controllers do also extend to datasets). In brief, personal data protection laws impose obligations on organisations in possession of or having control

145 *My Digital Lock Pte Ltd* [2018] SGPDPC 3 at [21].

146 See also *Google v Vidal-Hall* [2015] EWHC Civ 311, HC (England & Wales), on the nature of this new tort, and *Von Hannover v Germany* [2004] EMLR 379, ECtHR, for the genesis of the Court's recognition of the state obligation to secure privacy, where Arts 8 and 10 of the ECHR conflict with each other.

147 *Cf* definition of personal data in section 2 of the PDPA.

over personal data, while empowering individual data subjects with the ability to control how 'his or her' personal data is used and managed.

3.27 Therefore, it is appropriate to consider whether personal data protection laws, in particular the set of data subject rights provided by those laws, offer sufficient features (or 'incidents') of ownership to meet the expectations of individual data subjects in 'owning' or controlling their data.

3.28 One means of such assessment is to compare the protections offered by property rights with those offered by data subject rights in individual records. Such a comparison, in turn, requires a common benchmark. For this, we have looked to Tony Honoré's articulation of property rights as characterised by the following eleven key incidents: (1) the right to possess, (2) the right to use, (3) the right to manage, (4) the right to the income, (5) the right to the capital, (6) the right to security, (7) the incident of transmissibility, (8) the incident of absence of term, (9) the liability to execution, (10) the prohibition of harmful use, and (11) the residuary character.[148] While it is acknowledged that the inclusion of several of the aforementioned incidents is and has been subject to debate,[149] they provide a helpful starting point for our present analysis.

3.29 Singapore's data protection laws, set out in the PDPA, generally impose nine main obligations: (1) the consent obligation (section 13), (2) the purpose limitation obligation (section 18), (3) the notification obligation (section 20), (4) the access and correction obligations (sections 21 and 22), (5) the accuracy obligation (section 23), (6) the protection obligation (section 24), (7) the retention limitation obligation (section 25), (8) the transfer limitation obligation (section 26), and (9) the openness obligation (sections 11 and 12).[150] Given the impending introduction of data portability requirements, this obligation is also included in our analysis.

---

148 Tony Honoré, "Ownership" in Anthony Guest (ed), *Oxford Essays in Jurisprudence* (Oxford: Oxford University Press, 1961). See also Tang Hang Wu and Kelvin F K Low, *Tan Sook Yee's Principles of Singapore Land Law* (4th ed) (Singapore: LexisNexis, 2019), at 10.

149 See, for example, John Christman, *The Myth of Property: Toward an Egalitarian Theory of Ownership* (New York, NY: Oxford University Press, 1994) who argued that only the first five incidents in Honoré's conception are vital to ownership; Jeremy Waldron, *The Right to Private Property* (Oxford: Clarendon Press, 1990) who considered the prohibition on harmful use not to be an incident of ownership; and Alan Carter, *Philosophical Foundations of Property Rights* (London: Harvester Wheatsheaf, 1989) who concurred with Waldron and added that the liability to execution should not be regarded as an incident either.

150 Goh and Aw, above, n 140 at [4.14]; Tan Sin Liang, "How Well Do You Understand the Personal Data Protection Act and its Practical Implications?", *Singapore Law Gazette* (April 2014) <*http://v1.lawgazette.com.sg/2014-04/1014.htm*> (accessed 10 June 2020); and the PDPA, above, n 87, ss 11–26 and 32.

3.30 The extent to which those nine PDPA obligations in combination confer 'quasi-property rights' can be assessed – albeit somewhat crudely – by analysing whether they cover Honoré's 11 incidents of ownership, as the table below illustrates:

	Incident of ownership	PDPA obligation	Extent to which PDPA obligation maps on to incident / Comments
1	Right to possess	s 13 - the "Consent obligation"; s 20 - the "Notification obligation"	These Consent and Notification obligations are weakened, however, by the PDPA's "deemed consent" provision (s 13 & s 15) and the exceptions listed in Schedules 2, 3 and 4 PDPA.[151]
2	Right to use	As above.	As above.
3	Right to manage	s 13 - the "Consent obligation"; s 20 - the "Notification obligation"; s 21 - the "Access obligation"; and s 22 - the "Correction obligation"	The right to manage is "the right to decide how and by whom the thing owned shall be used" and depends on, among other things, the power "to permit others to use one's things [and] to define the limits of such permission"[152.] Access and correction obligations may be mapped to this incident of ownership as they concern the right of the data owner to determine how his data may be used by a third party (i.e. the data may be used by a third party with the conditions that it remains accessible to and available for correction by the data owner). The future data portability obligation will further strengthen individuals' right to manage their own personal data.
4	Right to the income	Not applicable.	PDPA does not prohibit data subjects from benefiting from the commercial exploitation of their personal data.
5	Right to the capital	Not applicable.	Honoré contemplates that the right to capital refers to such power to *"alienate the thing and the liberty to consume, waste or destroy the whole or part of it"*.[153] Such a concept may not be applicable to personal data.

151 See further paragraph 3.34 below regarding the conceptual challenges when considering a right to exclusive possession in the context of personal data.
152 Tony Honoré, "Ownership", above n. 148 at 372.
153 *Id.* at 373.

	Incident of ownership	PDPA obligation	Extent to which PDPA obligation maps on to incident / Comments
6	Right to security	Not applicable.	Honoré contemplates that a property owner is able to deal with his or her own asset in any manner that he or she deems fit, enjoying immunity from expropriation of the asset (apart from bankruptcy and execution for debt).[154] Such a concept may not be applicable to data.
7	Incident of transmissibility	Currently, the s21 access obligation allows the data subject to obtain a copy of his personal data that he can personally transmit.	The mapping of access obligations to the incident of transmissibility is somewhat tenuous, given that the crux of the obligation is not so much the data subject's capability to transmit his data but his right to access his personal data that is in the possession or under the control of a third party organisation. In future, data portability obligations will impose this obligation on the data controller to transmit at the request of the data subject. This obligation is more appropriately mapped to the incident of transmissibility.
8	Incident of absence of term	Not applicable.	Data subjects' rights under the PDPA are not subject to a fixed term.
9	Liability to execution	Not applicable.	Honoré contemplates that the liability to execution refers to the "liability of the owner's interest to be taken away from him for debt".[155] This concept is not applicable to personal data.
10	Prohibition of harmful use	Data subject may withdraw consent under s 16 of the PDPA if he becomes aware of harmful use of his personal data. This is reinforced by the overarching obligation to ensure that data is collected, used of and disclosed for a "reasonably appropriate purpose" under s 18 of the PDPA.	However, the "deemed consent" provision under s 13 & s 15, as well as lists of exceptions in Schedules 2, 3 and 4 of the PDPA weakens these obligations.

154 *Id*. at 375.
155 *Id*. at 374.

	Incident of ownership	PDPA obligation	Extent to which PDPA obligation maps on to incident / Comments
11	Residuary character	Not applicable.	In property law, when an interest less than ownership terminates (e.g. easements, leases), most legal systems will provide for rules to ensure that an owner can exercise the corresponding rights for those interests that have been terminated.[156] Such concepts are not applicable to personal data.

Table 1: Mapping the PDPA obligations to Honoré's Incidents of Ownership

3.31 As the table above demonstrates, there are some noticeable overlaps between the key incidents of ownership and the obligations under the PDPA. For example, the obligation to allow access to (and correction of) data and the proposed inclusion of the obligation to port data, correspond with the "right to manage" as an incident of ownership. Similarly, the PDPA consent and notification obligations are, to a degree, akin to a right (of the data subject) to possess the data in question. Elsewhere, overlaps are more tenuous, or gaps emerge. For example, it is not obvious that there is a PDPA counterpart for the ownership right to security (of the subject property).

3.32 It is acknowledged that establishing overlaps between the PDPA obligations and incidents of ownership is a somewhat crude measure for assessing the extent to which the PDPA provides for 'quasi-property' rights. With that caveat, however, we note that:

- While overlaps exist, the difficulty of mapping the two regimes across to one another (and the 'gaps' in coverage that remain when this is done) could be seen as indicative of the rather different nature of the rights accorded by data protection laws and property rights.
- Equally, however, it could be argued on closer examination that those gaps in coverage relate to aspects of property rights that may not be relevant for personal data:
    - For example, the 'missing' rights to capital and security are possibly inimical to personal data, since it will be impossible for an individual to alienate or assetise his biodata.
    - By comparison, the right to income is compatible with personal data and one that, albeit not explicitly a data subject right, is very much exercised *de facto*, e.g. giving consent for the collection of one's activity-generated

156 *Id.* at 375.

data is the *quid pro quo* for access to free online social media content.

3.33 The foregoing suggests that a bespoke property right may not be necessary if the objective is to confer on the data subject sufficient control over his personal data. Where the incidents of ownership make sense for personal data, existing or upcoming data subject rights already provide the ability to control how personal data is used. This calls into question whether it is necessary to provide greater protections over personal data than those that currently exist. It should also be borne in mind that personal data is unlike other forms of intangible property, and (as discussed in the preceding paragraph and traversed in the table above) there are incidents of ownership relevant to intangible or intellectual property that will be difficult to transpose to personal data.

2 Challenges with conferring ownership rights over personal data

3.34 There are also serious *conceptual* difficulties with elevating control over individual records of personal data to the level of ownership rights (at least, within Honoré's incidents of ownership framework):

- <u>Non-rivalrous nature of data:</u> Personal data has a non-rivalrous character. This means that data can be possessed and exploited differently by two or more persons without diminishing the value to either. The geolocation information of an individual, for example, is recorded by GPS sensors on her mobile phone, but can be collected and used by multiple apps installed on that device.

  The non-rivalrous nature of data sits at odds with traditional incidents of ownership such as rights to possess, use and manage to the exclusion of others. It was this disjunction that the United Kingdom's House of Lords Select Committee on Artificial Intelligence had in mind when it recently concluded that there were conceptual difficulties inherent in any discussion of 'data ownership', and that 'data control' provided a more useful framework for assessing the appropriate allocation of rights over data.[157]

- <u>Non-excludability of data:</u> Closely linked to that 'non-rivalrous' nature is the fact that data also has a non-excludable characteristic[158]. Personal data – as the name suggests – are facts for which the concept of exclusion is inapplicable.

  Take identification information and contact details as examples. Personal data protection laws provide data subjects with some control over how organisations may use such

157 *Lords Committee AI Report*, above, n 121 at para 62.
158 *Data Ownership Rights and Controls*, above, n 124 at 20.

information (e.g. no telemarketing messages), However, that control is not exclusive: as data protection law recognises various legitimate uses of that data (e.g. keeping track of customers who have opted out from receiving telemarketing messages), it falls some way short of a right to exclude.[159]

The 'right to possess' incident of ownership, by contrast, is predicated on such exclusivity of control. Even in the space of intellectual property, for example, patent owners are able to assert exclusive rights to the use of their patents through exclusive and/or non-exclusive licensing.

- <u>Inability to alienate the data:</u> The non-rivalrous and non-excludable character of personal data means that there will be inherent difficulties with the alienation of such data, a power which falls under property owners' 'right to the capital'.

  The crux of this problem is that the disclosure or transfer of an individual's personal data to a third party does not necessarily give that third party formal title to the information.[160] This is because the data provided by an individual may be "used without being used up" (i.e. non-rivalrous) and can be "sold without being relinquished" (i.e. non-excludable).[161] If data is disclosed or transferred to a third party, formal and exclusive title of that data can hardly be granted, unless the parties contract for exclusive rights to be given over the disclosure of the data.

- <u>Expansibility of data:</u> A further difficulty lies in the expansibility of data. While the benefit associated with a single record of data is miniscule on its own, there is tremendous economic and social impact when that record of data is aggregated, and combined with other records of data. Further secondary value of that single record of data can arise from the initial aggregation of data.

  In this regard, one would face difficulties quantifying the utility and value of a single record of data from a user by itself. While difficulties in valuation of individual records of data should not *ipso facto* be obstacles to conferring property rights, they add to the arguments against the significant changes in policy that would be required in order to confer such rights.

159 Even the right to erasure (or to be forgotten) as implemented in the EU GDPR is subject to a list of exceptions: Article 17(3).
160 Aplin, "Confidential Information as Property?" above, n 133.
161 RT Nimmer and PA Krauthaus, "Information as Property: Databases and Commercial Property'" (1993-1994) 1 International Journal of Law and Information Technology at 10–11.

3.35 It is important also to consider the *policy* challenges that conferring property rights of data may cause for a data-driven economy such as Singapore. In that regard, it is notable that, based on our review of several jurisdictions' experiences, no jurisdiction has to our knowledge elevated individual records of data to the status of a property right.

3.36 Insofar as Singapore is concerned, we have identified various policy risks, in particular:

- <u>Raising the barriers to entry over data:</u> Although it has been argued that property rights, and by extension ownership, may offer a viable solution to the problem of commodified (personal) data, this supposed solution is in itself problematic.

  Apart from difficulties in valuing individual records of personal data, according such rights would create a barrier to the big data phenomenon and the process of data mining, while impeding the realisation of their positive externalities (e.g. better internet services and processes).

  Corporations (particularly SMEs) may well feel that the cost of complying with an ownership regime outweighs any benefit that may be accrued by the eventual use of the collected big data. That might particularly be the case for uses that are more experimental or innovative, and thus less certain to result in commercial benefit – yet it is precisely that sort of innovation that policy in this area typically seeks to encourage.

- <u>Disruption to existing legal framework:</u> As established above, the existing legal frameworks do not recognise a property right to data or information. Instead, they each seek to protect a tangential incident of data and information (e.g. breach of confidence effectively guards against the wanton exploitation of confidential information) and thereby indirectly govern data and information. By recognising that there can be ownership of data and information (in the property sense), we risk disrupting the established norms in these and other areas of law (e.g. criminal laws on theft and misappropriation of property).

C RECOMMENDATIONS

3.37 Our review indicates that there may not be a strong normative impetus for shifting from an information custodian approach to a property rights regime with respect to individual records of data. In fact, commentators have cautioned that a property rights regime may not be as protective as it seems, and highlighted that the "rights, entitlements and the power to control need not necessarily be associated with ownership."[162]

3.38 In our view, the conceptual challenges of data's intangibility, combined with the doctrinal disruption such a regime would entail, render it ultimately incoherent, and thus undesirable. If conferring a particular right or entitlement over personal data was felt to be sufficiently important, it would be entirely possible to do so specifically through other legal means (e.g. legislation or common law). The impending introduction of the data portability obligation in the PDPA is one such example.

3.39 We note that the impetus for much of the debate regarding data ownership is the mishandling of data by technological organisations such as Cambridge Analytica. However, this mischief can also be addressed by implementing specific controls, rights and entitlements that can be exercised by an individual or organisation. The fact that data protection authorities have been able to impose sanctions under current data protection laws (powers which are, furthermore, being strengthened)[163] speaks to the adequacy of the present regimes.

3.40 In this respect, the PDPA recognises "both the right of individuals to protect their personal data and the need of organisations to collect, use or disclose personal data" for reasonably appropriate purposes.[164] As Table 1 above illustrates, there in considerable overlap between the incidents of ownership that are relevant to data and the statutory obligations and entitlements around personal data. More importantly, the PDPC has conducted several public consultations to review the PDPA, including introducing data portability obligations, and data innovation provisions. If the underlying object of considering a property rights regime to data is to ensure that individuals' privacy over that piece of data is respected, then we consider such an object can be achieved by ensuring that any valuable rights or entitlements identified are enshrined in the PDPA.

162 *Data Ownership Rights and Controls*, above, n 124 at 11.

163 Amendments to the PDPA have been proposed to increase the PDPC's fining powers to up to either 10% of a company's annual turnover or S$1 million, whichever is higher. At present, penalties are capped at S$1 million. *https://www.mci.gov.sg/public-consultations/public-consultation-items/public-consultation-on-the-draft-personal-data-protection-amendment-bill* (accessed 10 June 2020)

164 Section 3 of the PDPA.

CHAPTER 4

CONCLUSION

4.1 The question of ownership of data is not an easy one to answer, and requires consideration, in particular, of the differences between datasets or individual records of data. However, this dichotomy is not new, and the law has already had to deal with compilations and database rights (for datasets) and has established principles over when information (i.e. individual records of data) can be protected (e.g. confidential information, personal data, etc).

4.2 We considered whether law reform – and in particular legislative intervention – was necessary in relation to the laws protecting datasets, considering both legal and economic perspectives. Our conclusion is that clarifications as to how compilation rights apply for the copyright protection of electronic databases are desirable. However, we do not consider that that requires primary legislative reform. Rather, we believe that clarifications through administrative guidance from IPOS (or equivalent 'soft law' measure) would be sufficient and effective. Such guidance would also be helpful to overcome evidential challenges in the modern data entry and data governance workflow to ensure that records of authorship are properly kept.

4.3 While we acknowledge the gap that *sui generis* database rights filled in the EU context, and recognise its interplay with copyright compilation rights, we were not able to identify clear evidence that the introduction of such rights had a significant positive correlation with developments in the data economy. We were mindful also of the jurisprudential differences between EU and Singapore's intellectual property regimes, in particular the fact that the threshold of intellectual input required to meet 'originality' requirements in the EU appears higher than the 'sweat of the brow' threshold under Singapore's copyright regime. Thus, the conceptual gap that the EU *sui generis* right was designed to fill arguably does not exist to the same extent here.

4.4 In the final analysis, therefore we do not make any recommendations for reform in this regard. However, in the light of increasing automation, we do see benefit in copyright protection of computer-generated works being recognised through legislative amendments, with guidance being provided in the interim on when computer generated works enjoy copyright protection.

4.5 This report has also considered the protection of individual records of data. While there may be an instinctive desire to use ownership and property rights as legal frameworks to control data that we consider to be 'ours', there are fundamental difficulties – on grounds of jurisprudential principle and policy – to so doing.

4.6 These stem in part from individual records of data being in the nature of *facts*, rather than tangible assets or creative expressions. For example, copyright requires sufficient expression of the idea, while the law of confidentiality ceases to protect information that is publicly available. These branches of law evolved and continue to operate effectively to protect verbose information like a biography or a secret recipe, but were not intended to apply to collections of factual data points.

4.7 Data protection laws give data subjects the ability to control their *personal* data in the custody of organisations. Having compared the extent of control offered by data protection law against the principal incidents of property rights (as proposed by Tony Honoré), it is our conclusion that existing and incoming data protection laws currently provide data subjects with the ability to exercise an appropriate degree of control over their personal data. As such, we do not consider it profitable to replace this established framework with a new one constructed from property law.

4.8 There is no equivalent to the PDPA governing *non-personal* data. While the issues regarding such data are in certain respects less acute than for personal data, we see merit in consideration being given to whether a right akin to data portability should be introduced for non-personal data. For example, clarification that the user of a 'smart' device also has the right to gain access to and acquire a copy of the data generated by their use of the product could serve similar policy objectives to those underpinning the rights of access and portability over personal data. The benefits of such any such right may become more pronounced as – among other things – IoT sensors and advances in telecommunications networks continue to drive growth in the volume and variety of machine-generated data.

4.9 The basis for granting this right could be found in consumer protection law, following the model of Australia, where the newly-introduced consumer data right can be exercised not only by consumers but also by business entities acting as consumers. With the impending introduction of data portability obligations for personal data under the PDPA, we recommend that the implementation of such obligations and their effectiveness in achieving the stated policy objectives are monitored, before returning to the question of extending such a portability concept to non-personal data.

GLOSSARY[165]

AI system — a machine-based system able, for a given set of human-defined objectives, to make predictions, recommendations, or decisions that influence real or virtual environments. Such systems are able to operate with some level of **autonomy**, and can be incorporated into hardware devices or entirely software-based.

Algorithm — a set of rules or instructions (i.e. mathematical formulas and/or programming commands) given to a computer for it to complete a given task.

Application Programming Interfaces (APIs) — a set of instructions that enables interaction and integration between two or more software systems.

Artificial Intelligence (AI) — a set of technologies that seek to simulate human traits such as knowledge, reasoning, problem solving, perception, learning and planning, and, depending on the AI model, produce an output or decision (such as a prediction, recommendation, and/or classification).[166]

Auditability — the readiness of an **AI system** to undergo an assessment, by internal or external auditors, of its **algorithms**, **data** and design processes.

Autonomy/autonomous — the ability of an **AI system** to function (i.e. to take decisions and act) independently without human intervention.

Bias — the distortion or skewing of an **AI system**'s outputs, either due to the design of the algorithm or due to the input **datasets** utilised by the AI system being unrepresentative or discriminatory. Two common forms of bias in data include:

- selection bias (when the **data** on which an **AI system** bases its outputs are not representative of the actual **data** or environment in which the **AI system** operates); and
- measurement bias (when the process or means by which **data** is collected results in that gathered **data** being skewed or distorted).

165 The definitions in this glossary have been adapted from various sources for the specific purposes of the present series of reports. They are intended as an aid to the reader and should not be treated as exhaustive or authoritative.

166 We note that there is no widely-accepted or authoritative definition of artificial intelligence. The definition used here is a non-exhaustive, adapted definition used in the *Model AI Governance Framework (Second Edition)*, above, n 7.

Big Data — datasets characterised by their:

a) size (“Volume”);

b) complexity (“Variety”) (i.e. typically including structured, semi-structured and unstructured data derived from diverse sources); and/or,

c) rate of growth (“Velocity),

from which detailed insights can be derived using advanced analytical methods and technologies (e.g. **neural networks** and **deep learning**).

Black box (1) —an **AI system** whose decision-making operations are not **explainable** – that is, the means by which it reached a particular decision or action are neither disclosed nor able to be ascertained by human **users** or other interested parties (for example regulators, testers or auditors).

Black box (2) — see **Event Data Recorder**.

Bot — a software program (typically operating on the internet) designed to run automated tasks.

Chatbot — an **AI system**, commonly used in customer-facing commercial settings, designed to engage in dialogue with a human **user** via voice or written methods, and thus to simulate a human-to-human conversation. As the Chatbot engages in more conversations, it learns to better respond to future questions and more closely imitate real conversations. Examples include the “Ask Jamie” chatbot on the Singapore Ministry of Health’s website, or the ‘Live Chat’ help functions on e-commerce platforms such as Lazada or Shopee.

Cyberattacks — a malicious attack launched from one or more computers against other computers, networks or devices.

Data — information defined as and stored in code to be processed or analysed. Individual records of data (for example a person’s name or the temperature recorded by a smart home device at a particular date and time) can be combined together to form **datasets**. A distinction is commonly drawn between personal data (those which individually or in combination with other data, identify an individual) and non-personal data (those that do not).

Data portability — the legal obligation to comply with a data subject’s request for their **data** to be moved from one organisation to another in a commonly used machine-readable format.

Database management system (DBMS) — software that enables users to create, update, retrieve, and manage **data** within a database using defined commands.

Dataset — a collection of **data** (often stored in the form of one or more databases).

Deep learning — a specific form of **machine learning** that utilises **neural networks** to model and draw insights from complex structures and relationships between **data** and **datasets**. The term derives from the 'layers' of the **neural network** down through which the **data** passes.

Deployer — the person or legal entity responsible for putting an **AI system** on the market or otherwise making it available to users. The deployer may also have on ongoing role in operating or managing the **AI system** after deployment.

Derived data — any **data** element that is created and/or derived by an organisation through the processing of other **data** in the possession and/or control of the organisation.

Designer / Developer — a person or legal entity who takes decisions that determine and control the course or manner of the development of **AI systems** and related technologies. 'Development' for these purposes means a) designing and constructing **algorithms**, b) writing and designing software, and/or c) collecting, storing and managing **data** for use in creating or training **AI systems**.

Event Data Recorder — a machine that continuously records the inputs received by an **AI system** (e.g. what its sensors 'see'), its relevant internal status data, and its outputs. Sometimes colloquially known as a 'black box recorder'. The intention of such event data recorders, equivalent to those installed in aircraft, is to allow post-hoc analysis of the **AI system**'s operation (e.g. in the lead up to an accident or system failure).

Explainability — the ability for a human, by analysing an **AI system**, to understand how and why the system reached a particular decision or output.

Explainable AI — broadly, either a) AI systems which are designed to be inherently **explainable**, such that a human can understand how and why the system reached a particular decision or output; or b) tools designed to help extract explanation from pre-existing **black box** and other complex **AI systems**.

Human-Machine Interface — a screen, dashboard or other interface which enables a human **user** to engage with an **AI system** or other machine.

Internet of Things, the (IoT) — a system comprised of interconnected devices (commonly known as smart devices) that transfer **data** and communicate with one another via the internet.

Machine Learning — a technique whereby a set of **algorithms** utilise input **data** to make decisions or predictions, and thus to 'learn' how to complete a task without having been specifically programmed to do so.

(Artificial) Neural Networks — a series of 'layered' **algorithms** used to analyse, classify, learn from and interpret input **data**. The values from one layer are fed into the next layer to derive increasingly refined insights. Artificial Neural Networks are so named because they broadly mimic the biological neural networks in the human brain.

Operator — see **User**.

Robotics — technologies that enable machines to perform tasks traditionally performed by humans, including by way of **AI** or other related technologies. This series of reports focuses on robots that act fully or partially autonomously, without human intervention.

Robustness — the ability of an **AI system** to deal with errors that arise during execution or erroneous input, and to continue to function as intended or without insensible, unexpected or potentially harmful results.

Structured data — data that is highly-organised and formatted according to pre-defined fields (for example a table listing individuals' names in columns alongside their height and weight), making it simpler to search and analyse.

Three-tier architecture — in the context of application architecture, the functional or physical separation of the database's data management, application/input processing and visual presentation functions. Those separated functions are commonly termed, respectively, the data layer, the application or logic layer and the presentation layer.

Traceability — the documentation, in an easily understandable way, of (a) an **AI system**'s decisions, and (b) the **datasets** and processes that yield those decisions (including those of data gathering, data labelling and the **algorithms** used). This provides a means to verify the history, and contexts in which decisions are made.

Transparency — various mechanisms or requirements intended to provide additional information to users, regulators and other stakeholders regarding the algorithmic decision-making processes undertaken by **AI systems**, and the input **data** relied on by such systems. Such transparency may be achieved through, for example, disclosure of source code, **explainability** and/or **traceability**. Transparency also implies that **AI systems** should (in practice, and by design) carry out their functions in the way communicated to others (including **users**).

Unstructured data — data that is not organised, formatted or tagged in any pre-defined way (for example images or audio files, or free text email content), and thus is harder to search and analyse.

User — any natural or legal person who uses an **AI system** for purposes other than development or deployment.

Verifiability — the process of ensuring that the outputs of an **AI system** correspond with its intended function or purpose (for example by testing the system using a range of different inputs, or ensuring that a particular input consistently and repeatedly leads to a desired output).

ISBN 978-981-14-6597-0 (softcover)
978-981-14-6598-7 (e-book)