COPYING RIGHT IN COPYRIGHT LAW

Fair Use, Computational Data Analysis and the Personal Data Protection Act

Data mining and predictive analytics is a multi-billion industry in the 21st century and can generate immense public benefit. At the same time, it has attracted global controversy in highly visible cases of copyright and data protection violations such as in the Google Books and Google Images litigation. The governing laws require a constant delicate balancing between the interests of rights owners, data miners and private individuals. This article analyses Singapore's copyright and data protection regimes to examine their effectiveness as dual and complementary regulators of data mining. It also identifies the salient concerns shared by stakeholders to pinpoint avenues for legal reform. Finally, in recognising that economic considerations play a crucial role underlying Singapore's copyright and data protection regimes, this article suggests that an economic analysis to identify reforms in line with maximising welfare among stakeholders is apropos. The article concludes that more illustrations should be added to the proposed computational data analysis exception under s 244 of the proposed Copyright Act 2021 to clarify the grey areas, and that the research exception under the Personal Data Protection Act 2012 (Act 26 of 2012) should be retained in its current form.

David TAN¹

PhD (Melbourne), LLM (Harvard), LLB (Hons), BCom (Melbourne); Professor, Faculty of Law, National University of Singapore.

Thomas LEE Chee Seng¹

LLB (Hons) (National University of Singapore), BA (Hons) (Yale-NUS).

© 2021 Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

¹ The authors would like to express their appreciation to Benjamin Wong and the anonymous reviewer for their comments on an earlier draft of this article.

I. Introduction

1 In 1989, British Computer Scientist Tim Berners-Lee, frustrated with the inefficiencies of finding information stored on different computers, invented the World Wide Web.² Almost 60% of the global population are estimated to be active Internet users today,3 with the amount of data uploaded to the Internet amounting to a staggering 24,000 gigabytes per second.⁴ With this explosively growing body of online data, powerful tools were desperately needed to uncover valuable information and transform such data into organised and understandable knowledge. This necessity led to the birth of data mining ("DM"), and the associated development in predictive analytics, powered by advances in artificial intelligence ("AI") and powerful computers that provided the infrastructure necessary for the systemic development of DM tools that now turn big data into golden nuggets of knowledge.⁵ The predictive analytics market is estimated to grow at a compound annual growth rate ("CAGR") of 24.5% and reach US\$22.1bn by 2026.6 Predictive analytics a category of data analytics - are increasingly used to obtain information from historical and current datasets using statistical modelling and machine learning techniques to forecast potential future findings and trends.

² Ashley May, "Happy 30th Birthday, World Wide Web. Inventor Outlines Plan to Combat Hacking, Hate Speech" USA Today (12 March 2019) (accessed 11 September 2021).

³ Joseph Johnson, "Global Digital Population as of January 2021" *Statista* (10 September 2021) https://www.statista.com/statistics/617136/digital-population-worldwide/ (accessed 11 September 2021).

^{4 &}quot;The Internet in Real Time: Web Usage Stats Per Second" *WebFX* (17 May 2021) https://www.webfx.com/internet-real-time/> (accessed 11 September 2021).

⁵ Han Jiawei, Micheline Kamber & Jian Pei, *Data Mining: Concepts and Techniques* (Morgan Kaufmann, 3rd Ed, 2012) at p 5.

^{6 &}quot;At 24.5% CAGR, Global Predictive Analytics Market Size to Register Record Value of USD 5.7 Billion by 2026, Says Facts & Factors" *GlobeNewswire* (18 March 2021) <https://www.globenewswire.com/news-release/2021/03/18/2195402/0/en/At-24-5-CAGR-Global-Predictive-Analytics-Market-Size-to-Register-Record-Value-of-USD-5-7-Billion-by-2026-Says-Facts-Factors.html> (accessed 17 May 2021). See also Bernard Marr, "The 4 Biggest Trends in Big Data and Analytics Right for 2021" *Forbes* (22 February 2021) <https://www.forbes.com/sites/bernardmarr/2021/02/22/the-4-biggest-trends-in-big-data-and-analytics-right-for-2021/?sh=7cac0b947df8> (accessed 17 May 2021) and John Edwards, "What is Predictive Analytics? Transforming Data into Future Insights" *CIO* (16 August 2019) <https://www.cio.com/article/3273114/what-is-predictive-analytics-transforming-data-into-future-insights.html> (accessed 17 May 2021).

No part of this document may be reproduced without permission from the copyright holders.

Three decades after the founding of the World Wide Web, the 2 world's biggest social network Facebook found itself at the centre of an international scandal involving voter data. Consultants working for Donald Trump's presidential campaign exploited the personal data of millions of Facebook users. Cambridge Analytica, a data mining firm, had reportedly acquired Facebook user data in a way that violated Facebook's policies by building psychographic profiles of users and their friends for targeted political advertisements.⁷ More recently in 2020, a Wall Street Journal investigation found that Amazon was mining data from third-party sellers to help develop its private-label goods, despite assurances to Congress to the contrary.8 This has reportedly helped them to make decisions on whether they should enter a new product category.9 What these cases illustrate is an increasing trend by firms to apply DM techniques for contentious purposes that have not yet been adequately addressed in the law by both regulators and judges. In a country like Singapore that strongly embraces the digital future,¹⁰ this raises challenging legal questions.

3 Understandably, the bulk of improper DM activities falls under the aegis of data protection laws, with hefty fines often imposed on errant corporations. In 2019, the US Federal Trade Commission voted to approve a fine of US\$5bn against Facebook on the grounds of user-privacy violations.¹¹ More recently in 2020, the UK Information Commissioner fined Ticketmaster £1.25m under the General Data Protection Regulation for a personal data breach, wherein a malicious chatbot skimmed credit card data from Ticketmaster's clients.¹² But increasingly, data and text

⁷ Ian Sherr, "Facebook, Cambridge Analytica, Data Mining and Trump: What You Need to Know" CNET (18 April 2018) https://www.cnet.com/news/facebookcambridge-analytica-data-mining-and-trump-what-you-need-to-know/> (accessed 17 May 2021).

⁸ Dana Mattioli, "Amazon Scooped Up Data From Its Own Sellers to Launch Competing Products" Wall Street Journal (23 April 2020) https://www.wsj.com/articles/amazon-scooped-up-data-from-its-own-sellers-to-launch-competing-products-11587650015> (accessed 17 May 2021).

⁹ Tyler Lee, "Amazon is Reportedly Mining Third-Party Seller Data to Create its Own Products" *übergizmo* (23 April 2020) https://www.ubergizmo.com/2020/04/amazon-third-party-seller-data-create-own-products/> (accessed 17 May 2021).

¹⁰ Royston Sim, "Right Mindset Needed to Embrace Digital Future: PM Lee" *The Straits Times* (9 July 2017).

^{11 &}quot;FTC to fine Facebook about \$5 Billion Penalty for User-Privacy Violations, Reports Say" CBS News (15 July 2019) https://www.cbsnews.com/news/ftc-to-fine-facebook-5-billion-for-cambridge-analytica-user-privacy-bungles-reports/ (accessed 11 September 2021).

¹² Ruth Boardman, "UK: Analysis of Information Commissioner's £1.25 Million Fine for Ticketmaster Personal Data Breach" Bird & Bird (1 November 2020) http://www.twobirds.com/en/news/articles/2020/uk/analysis-of-information-(cont'd on the next page)

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

mining involves the copying, digitisation or reformatting of copyright material which may give rise to copyright infringement.

However, no DM case has been litigated in Singapore to date. 4 Although Singapore revised its data protection laws by amending the Personal Data Protection Act 2012¹³ ("PDPA") in 2020,¹⁴ a more direct response to regulate DM activities is found in the 2019 Copyright Review Report¹⁵ and the Copyright Bill.¹⁶ The Copyright Bill, released for public consultation on 5 February 2021, sought to introduce a specific exception to copyright infringement for DM which is phrased as a specific exception for "computational data analysis" under a new s 244. Computational data analysis is defined under s 243(a) as the use of a computer program to "identify, extract and analyse information or data from the work" which is synonymous with text and data mining - and miners must prove that they have *lawful access* to online works or data in order to qualify for protection under this provision.¹⁷ While laudable, these legal regimes may have certain flaws, and there is significant uncertainty surrounding the proposed computational data analysis exception.

5 We will be evaluating whether Singapore's copyright and data protection regimes fare well in regulating DM activities. Firstly, while there is growing academic literature discussing doctrinal issues within copyright and data protection regimes, scholarship is lacking in how such issues may arise when applied to DM. Secondly, it is equally important that such academic concerns, if identified, should be shared by DM stakeholders in Singapore; otherwise these would be merely pedantic attempts to find fault where there is *practically* none. Thirdly, unlike the Hegelian and Kantian schools of copyright thought that have exerted a predominant influence in civil jurisdictions,¹⁸ copyright law in Commonwealth common law countries like Singapore has always

commissioners-1-25-million-fine-for-ticketmaster-personal-data-breach> (accessed 11 September 2021).

¹³ Act 26 of 2012.

¹⁴ Personal Data Protection (Amendment) Act 2020 (Act 40 of 2020).

¹⁵ Ministry of Law and Intellectual Property Office of Singapore, Singapore Copyright Review Report (17 January 2019) https://www.mlaw.gov.sg/files/news/pressreleases/2019/01/Annex%20A%20-%20Copyright%20Review%20Report%20 16%20Jan%202019.pdf> (accessed 17 May 2021).

¹⁶ Bill 17 of 2021.

¹⁷ Copyright Bill (Bill 17 of 2021) s 244(2)(d). It appears that the requirement of "lawful access" under Proposal 8 of the Copyright Review Report has been transposed to s 244(2)(d) of the Copyright Bill as one of its new requirements, in addition to other requirements under s 244(2) of the Copyright Bill.

¹⁸ Neil Netanel, "Alienability Restrictions and the Enhancement of Author Autonomy in United States and Continental Copyright Law" (1994) 5 Cardozo Arts & Entertainment Law Journal 12.

No part of this document may be reproduced without permission from the copyright holders.

been underpinned by economic rationales.¹⁹ Similarly, Singapore's data protection regime is not premised on a right of privacy, but recognises data as a "key economic asset" and the value that data analytics provides to "inform decisions [and] generate efficiencies".²⁰ Any potential legal reforms in Singapore must hence prioritise economic considerations to ensure that the aggregate welfare among DM stakeholders can continue to be maximised.

This article adopts a combination of doctrinal, empirical and 6 economic analyses to examine whether the current copyright and data protection regimes are rightly poised to regulate DM in Singapore. Together, they inform the normative perspective, which investigates how DM ought to be regulated. Part II²¹ sets out an overview of DM and the different types of stakeholders involved. Part III²² explains our overarching research thesis and methodology. Part IV²³ evaluates the current state of copyright and personal data protection laws in Singapore. Part V²⁴ presents the qualitative findings from our interviews with a range of data owners and miners. Part VI²⁵ considers potential solutions from an economic analysis perspective by comparing cost differentials between different methods of legal reform. Part VII²⁶ summarises key findings and advances some suggestions to clarify ambiguities in the proposed computational data analysis exception. Part VIII27 concludes with a reminder that with predictive analytics on the rise, one needs to ensure that a fair balance is properly struck between the remuneration for authors, and the access that should be granted to other users to enable them to copy these works in order to create new ones in promoting the public good.

- 22 See paras 14-19 below.
- 23 See paras 20–55 below.
- 24 See paras 56–64 below.
- 25 See paras 65–71 below.
- 26 See paras 72–85 below.
- 27 See paras 86–89 below.

© 2021 Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

¹⁹ Parliamentary Debates, Official Report (16 November 2004), vol 78 at col 1053 (S Jayakumar, Deputy Prime Minister and Minister for Law): "a sound IP regime is absolutely important for *future economic growth*" [emphasis added].

²⁰ *Parliamentary Debates, Official Report* (2 November 2020), vol 95 (S Iswaran, Minister for Communications and Information and Minister-in-charge of Trade Relations).

²¹ See paras 7-13 below.

II. Overview of text and data mining

A. Definition of data mining

7 DM, a term first coined in 1989,²⁸ is defined in the Copyright Review Report as "the use of automated techniques to analyse text, data, and other content to generate insights and information that may not have been possible to obtain through manual effort".²⁹ The term "text and data mining" is used in the report, but this will be collectively referred to as DM in this article, since the broader meaning of "data" encompasses "text". With this broad definition, DM applied to Internet data would understandably encompass *web mining*, which refers to the use of web robots to collect data from the web servers to discover patterns.³⁰ These web robots are usually open source in nature, which ranges from crawlers such as Heritrix³¹ to scrapers such as Scrapy.³²

8 The process of DM generally follows a number of stages. Raw data or works are first collected, before filtering methods extract useful data (*ie*, target data). Target data is then refined (this includes removing outlier data or detecting missing data) to generate preprocessed data. At the transformation stage, data is then consolidated into forms appropriate for mining, such as through normalisation³³ or discretisation³⁴ techniques.³⁵ Finally, algorithms are further applied to transformed data in order to obtain patterns, from which the user will apply the user's own interpretation to derive knowledge from those patterns. Figure 1 presents the above stages in the form of a flowchart.

 $^{\odot}$ 2021 Contributor(s) and Singapore Academy of Law.

²⁸ Yanli Cai & Jian-Tao Sun, "Text Mining" in *Encyclopedia of Database Systems* (Ling Liu & M Tamer Özsu gen eds) (Springer, 2009) at pp 3061–3065.

²⁹ Ministry of Law and Intellectual Property Office of Singapore, *Singapore Copyright Review Report* (17 January 2019) at p 32.

³⁰ Xiaoling Shu, *Knowledge Discovery in the Social Sciences: A Data Mining Approach* (University of California Press, 2020) at p 210.

³¹ Heritrix website https://webarchive.jira.com/wiki/spaces/Heritrix/overview (accessed 15 December 2020). Notably, Heritrix is designed to respect the robots.txt exclusion directives.

^{32 &}quot;Making Web Crawlers Using Scrapy for Python" *Datacamp* (12 January 2019) <https://www.datacamp.com/community/tutorials/making-web-crawlers-scrapypython> (accessed 15 December 2020).

³³ Normalisation is the process when data is scaled so as to fall within a smaller range, such as 0.0 to 1.0.

³⁴ Discretisation is the process when the raw values of a numeric attribute (*eg*, age) are replaced by interval labels (*eg*, 0–10, 11–20, *etc*) or conceptual labels (*eg*, youth, adult, senior).

³⁵ Han Jiawei, Micheline Kamber & Jian Pei, *Data Mining: Concepts and Techniques* (Morgan Kaufmann, 3rd Ed, 2012) at p 112.

No part of this document may be reproduced without permission from the copyright holders.



Figure 1: Stages of data mining³⁶

B. Text and data mining ecosystem

⁹ The reach of DM technology today encompasses the entire digital infrastructure on which data can be found, and its impacts on our daily lives are ubiquitous.³⁷ Electronic storage of data has opened a new gate to DM by making data retrieval and archival easier.³⁸ The open-source nature of web robots has also lowered any barriers to adoption of DM technology.

10 In line with its readily available technological infrastructure, DM is driven by various stakeholders comprising mainly of data owners ("owners")³⁹ and data miners ("miners"). To set in motion DM, owners must first make their data available on the Internet before miners can mine/collect data for analysis. In turn, DM can then be employed for commercial reasons (*eg*, companies can adopt DM to tailor promotions to customer profiles for targeted advertising⁴⁰), or non-commercial purposes (*eg*, academic research).

11 For certain owners who wish to commercially exploit their data (which may reside in works protected by copyright), they may employ technological means to prevent unauthorised access to their data,

³⁶ Thales Sehn Korting, "How Data Mining Works" (24 November 2015) https://prezi.com/v2zmhstglmoi/how-data-mining-works/?utm_campaign=share&utm_medium=copy> (accessed 16 March 2021).

³⁷ Han Jiawei, Micheline Kamber & Jian Pei, *Data Mining: Concepts and Techniques* (Morgan Kaufmann, 3rd Ed, 2012) at p 618.

³⁸ H A Al-Odan & A A Al-Daraiseh, "Open Source Data Mining Tools", 2015 International Conference on Electrical and Information Technologies at pp 369–374.

³⁹ For the purposes of this article, "owners" will be used as a broader term encompassing rights-holders whose work attracts copyright protection.

⁴⁰ Han Jiawei, Micheline Kamber & Jian Pei, *Data Mining: Concepts and Techniques* (Morgan Kaufmann, 3rd Ed, 2012) at p 619.

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

ranging from paywalls to digital encryption.⁴¹ Another frequent method is the adoption of terms of service ("TOS") agreements displayed on webpages that define the acceptable use of the site.⁴² These agreements typically appear as "click-wrap" agreements that require the activation of an "I agree" button, or "browse-wrap" agreements that are provided as a hyperlink on the webpage for viewing.⁴³ Finally, owners may even adopt a Robots Exclusion Protocol ("REP") – a standard ubiquitously used by websites that directly communicates with web crawlers or scrappers to prohibit them from processing certain areas of a website, which is usually packaged in the form of robots.txt files. Figure 2 shows how REPs can be located at the root directory of certain web servers such as Google.⁴⁴



Figure 2: Screenshot of a robots.txt file on Google's webpage, which specifies areas that are available for ("Allow") or restricted from ("Disallow") automated interaction

⁴¹ Gove N Allen, Dan L Burk & Gordon B Davis, "Academic Data Collection in Electronic Environments: Defining Acceptable Use of Internet Resources" (2006) 30(3) *Management Information Systems Quarterly* 601.

⁴² Gove N Allen, Dan L Burk & Gordon B Davis, "Academic Data Collection in Electronic Environments: Defining Acceptable Use of Internet Resources" (2006) 30(3) *Management Information Systems Quarterly* 601.

⁴³ Eliza Mik, "Contracts Governing the Use of Websites" [2016] Sing JLS 73. See for instance webpages like www.datarobot.com that displays a "browse-wrap" agreement on https://www.datarobot.com/terms-of-service/, which grants endusers a non-transferable license to "access and make use of the website", but prohibits "any use of data mining, robots, or similar data gathering and extraction tools".

⁴⁴ Gove N Allen, Dan L Burk & Gordon B Davis, "Academic Data Collection in Electronic Environments: Defining Acceptable Use of Internet Resources" (2006) 30(3) Management Information Systems Quarterly 601 at 609.

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

12 However, such technological means alone may be insufficient to prevent unauthorised DM since it is possible to circumvent the technology such as by cracking the encryption.⁴⁵ They may also be impractical due to increased demands on the digital infrastructure.⁴⁶ In addition, REPs are purely advisory and rely on the compliance of the web robot, which runs the risk of malicious web robots that are unlikely to honour them.⁴⁷ Instead, owners may rely on legal remedies, such as court injunctions or the award of damages, to deter miners from inappropriately using their data. Nonetheless, it remains to be seen in the Singapore context whether legal avenues would be widely adopted by owners to enforce their rights.

13 A hurdle in the way of legal remedies is that the jurisprudential development of DM within applicable legal regimes like copyright and data protection is entirely absent given that courts in Singapore have not had the occasion to consider this issue. In the face of these uncertainties, miners are also concerned that deep-pocket owners would still rely on such legal mechanisms to chill DM activities, and the costs of responding to a cease-and-desist letter or defending a lawsuit would be substantial for start-up firms. In addressing these myriad problems and to properly balance the different interests at stake, any revision to Singapore laws must also ensure promotion of economic efficiency given economic considerations that underlie the copyright and data protection regimes in Singapore.

⁴⁵ Gove N Allen, Dan L Burk & Gordon B Davis, "Academic Data Collection in Electronic Environments: Defining Acceptable Use of Internet Resources" (2006) 30(3) *Management Information Systems Quarterly* 601 at 609.

⁴⁶ Gove N Allen, Dan L Burk & Gordon B Davis, "Academic Data Collection in Electronic Environments: Defining Acceptable Use of Internet Resources" (2006) 30(3) Management Information Systems Quarterly 601 at 602.

^{47 &}quot;Robots.Txt Introduction & Guide" *Google Search Central* (3 January 2021) <https://developers.google.com/search/docs/advanced/robots/intro?visit_ id=637497792308429981-2177496564&rd=1> (accessed 3 January 2021).

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

III. Research methodology

14 Fundamentally, the goal of our research is to identify how one may more objectively strike the appropriate balance between the interests of owners and miners, and to assess normatively what the law *ought to be*. To achieve this end, this study adopts a three-part inquiry – (i) a doctrinal analysis of copyright and data protection laws; (ii) qualitative interviews with DM stakeholders in Singapore (that comprises a total of 22 owners and miners); and (iii) an economic analysis comparing possible reforms to the law.

15 This study recruited prospective respondents through snowball sampling.⁴⁸ They were then purposefully sampled to ensure that they are familiar with both DM technologies and the applicable legal regimes, such as copyright and data protection law, before being interviewed. Enrolled respondents that eventually fit these criteria ranged from academics to those with work experience in the private sector. The authors identified two samples of DM stakeholders comprising 11 miners (Category A) and 11 owners (Category B) each. The list of respondents is in Annex A.

16 For Category A, respondents were asked about their experience with DM technologies and questions regarding what they thought of the legal uncertainties in the current Copyright Act⁴⁹ (as well as the proposed DM exception) and the PDPA. Similar questions were posed to respondents in Category B to understand how their concerns might differ from those in Category A from the perspective of a rights holder and a potential claimant. The questions are listed in Annex B. The sampling and interview process is illustrated in Figure 3.

^{48 &}quot;Snowball Sampling" <http://changingminds.org/explanations/research/sampling/ snowball_sampling.htm> (accessed 14 February 2021). This is a non-probability sampling technique where existing study subjects recruit future subject from among their acquaintances.

⁴⁹ Cap 63, 2006 Rev Ed.

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.



Figure 3: Sampling and interview process

17 Given that Singapore's copyright and data protection regimes are fundamentally based on economic considerations, we later adopt an economic analysis to compare the projected *error* costs (*ie*, costs incurred when judges apply the law wrongly) associated with an open-ended fair dealing defence with that of a specific exception. This is because based on the reported cases after the 2004 amendments to the Copyright Act,⁵⁰ litigants have always relied on the open-ended fair dealing defence (the

⁵⁰ Notably, the Copyright (Amendment) Act 2004 (Act 52 of 2004) expanded the guidelines in s 35(2) to apply to "any purpose" except for a purpose of criticism or review (s 36) or a purpose of reporting current events (s 37). This contrasts with the state of the law prior to the amendments in 2004, where s 35(1) of the Copyright Act (Cap 63, 1999 Rev Ed) only narrowly provided for the defence of fair dealing *specifically* "for the purpose of research or private study" with the same multi-factorial approach under s 35(2) to determine whether such a dealing was fair. Accordingly, given how different the open-ended fair dealing defence under s 35(2) was as compared to prior to the 2004 amendments, any analysis of whether litigants chose to rely on the open-ended defence under the current s 35, or the specific exceptions under s 36 or 37, or both, *must* take reference from cases that are reported after the 2004 amendments that were gazetted on 1 January 2005.

current s 35(2)), but never the specific defence,⁵¹ which suggest that the relevance of a specific exception may have been rendered otiose since the inception of the open-ended fair dealing defence in 2004. Seen in the backdrop of the current Copyright Act that does not deprive the autonomy of litigators to adopt a hybrid approach (*ie*, pleading both defences), the lack of case law suggesting that a hybrid defence strategy was ever used thus calls into question the relevance of introducing a new specific exception at all, when mounting an open-ended fair use defence (s 190 of the proposed Copyright Act 2021) could possibly suffice in any given situation.

18 Moreover, even considering the fact that the proposed Copyright Act 2021 will soon *explicitly* allow such defences to be pleaded together,⁵² in practice, litigants may still inevitably find themselves having to choose only one due to the frequent incompatibility of both defences. The miner who seeks to justify that the DM activity was "transformative" under the open-ended fair use defence, thus obviating the need to seek an *ex ante* licence, could potentially undermine its pleadings (in the alternative under the computational data analysis exception) that the miner had "lawful access" to the original material. Take for example, the creation of a search engine: works are likely to have been copied *without lawful access* in order to create a full-text searchable database, but this would nonetheless have been a highly transformative purpose. In both *Authors Guild v Google, Inc*⁵³ and *Authors Guild, Inc v HathiTrust*⁵⁴ ("*HathiTrust*"), the US Second Circuit Court of Appeals found fair use, notwithstanding

⁵¹ Thus far, in all three reported cases after the 2004 amendments, litigants preferred to plead the open-ended defence in lieu of the specific fair dealing exception. In *RecordTV Pte Ltd v MediaCorp TV Singapore Pte Ltd* [2010] 2 SLR 152, the defendants tried to plead the open-ended defence under s 109 of the Copyright Act (Cap 63, 2006 Rev Ed), which is *in pari materia* to s 35 of the Copyright Act (s 109 provides the same fair dealing factors under s 35 but for audio-visual items as opposed to literary, dramatic, musical, or artistic works). No specific fair dealing exception was pleaded in the alternative. In *Public Prosecutor v St Hua Private School Pte Ltd, Song Chunwei* [2014] SGDC 342, the defendants only raised an open-ended fair dealing defence under s 35(2), which was eventually rejected by the District Court. No specific fair dealing defence was pleaded in the alternative. In *Global Yellow Pages Ltd v Promedia Directories Pte Ltd* [2016] 2 SLR 165, the defendant sought to rely on the open-ended fair dealing defence under s 35(2) but did not raise any specific fair dealing exception in the alternative.

⁵² Section 184 of the Copyright Bill (Bill 17 of 2021) reads: "Unless this Act expressly provides otherwise, a permitted use is independent of, and does not affect the application of, any other permitted use". Theoretically, this clarifies that a potential defendant should not be faced with having to make a choice between the openended defence and specific exception, which may presently be the case given the ambiguity inherent in the current version of the Copyright Act.

⁵³ Authors Guild v Google, Inc 804 F 3d 202 (2nd Cir, 2015).

⁵⁴ Authors Guild, Inc v HathiTrust 755 F 3d 87 (2nd Cir, 2014).

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

that the libraries had downloaded and stored complete digital copies of entire books, because such copying was essential to permit searchers to identify and locate the books in which words or phrases of interest to them appeared.⁵⁵ It would appear that an *un*-lawful access could be fair use, but would not otherwise qualify for the computational data analysis exception.⁵⁶ Thus, in Singapore, a hybrid legal defence strategy, even if theoretically plausible, is unlikely to be pleaded in practice especially if permission was not obtained prior to accessing the works. Instead, a miner is more likely to plead the fair use defence under the proposed ss 190 to 191 with a preponderance of US decisions in its favour. Recognising this reality of having to choose one defence or the other, an economic inquiry would nevertheless be useful to investigate which approach is favourable from an economic standpoint in the long run to related stakeholders.

¹⁹ Finally, to address the limitations under the research exception in the PDPA,⁵⁷ we investigated how the retention or abolishment of the consent obligation during data collection would affect costs incurred to DM stakeholders, as an additional factor to consider before deciding *whether*, and if so, *how*, the law should be reformed.

IV. State of copyright and data protection laws

A. Copyright

(1) *Copyright subsistence*

As a preliminary point, the Copyright Act in Singapore presently extends copyright protection only to original literary, dramatic, musical, or artistic works ("LDMA works"), and to specific subject matter other than works.⁵⁸ Generally, for LDMA works that are published online, the first publication of the work must take place in Singapore for copyright

⁵⁵ See Authors Guild v Google, Inc 804 F 3d 202 at 216–217 (2nd Cir, 2015) and Authors Guild, Inc v HathiTrust 755 F 3d 87 at 97–105 (2nd Cir, 2014).

⁵⁶ However, there are still numerous conditions under the proposed s 244(2) to be satisfied. It should be noted that He Tianxiang, who was comparing the approaches in East Asian jurisdictions to DM exceptions, proposed a less onerous exception for China phrased as "reproduction ... in the course of data analysis and mining, in order to uncover new knowledge or insights". Tianxiang He, "Copyright Exceptions Reform and AI Data Analysis in China: A Modest Proposal" in *Artificial Intelligence and Intellectual Property* (Jyh-An Lee, Reto M Hilty & Kung-Chung Liu gen eds) (Oxford University Press, 2021) ch 9, at pp 196–218.

⁵⁷ The research exception can be found under both Pts 2 and 3 of the Second Schedule to the Personal Data Protection Act 2012 (Act 26 of 2012).

⁵⁸ It is noted that the proposed Copyright Bill 2021 (Bill 17 of 2021) employs an umbrella term "work" to cover "authorial work" (traditionally known as literary, (cont'd on the next page)

No part of this document may be reproduced without permission from the copyright holders.

protection to still apply.⁵⁹ This is contrasted to bare facts that "fall outside the protection of copyright law".60 It is well-established in copyright law that facts, whether alone or as part of a compilation, are not original and therefore may not be copyrighted. Over a quarter century ago, the US Supreme Court delivered a landmark ruling in *Feist Publications*, Inc v Rural Telephone Service Co, Inc61 ("Feist") that rejected the "sweat of the brow" doctrine,⁶² and held that a factual compilation may receive copyright protection only if it featured an original selection or arrangement.⁶³ As a result of *Feist*, a preponderance of decisions, especially at the US Circuit Court level, have found factual compilations - for example, those which employed formats of compilations like telephone directories and horse racing guides which is not different from the convention in the industry - to be lacking in originality and therefore incapable of attracting copyright protection. Sitting as a rare five-member full bench in 2017, the Singapore Court of Appeal in Global Yellow Pages v Promedia Directories Pte Ltd⁶⁴ ("Global Yellow Pages") unanimously rejected the "sweat of the brow" doctrine in favour of the "creativity approach" in the context of compilations,⁶⁵ such that there must be a creation that is connected with "the application of intellectual effort, creativity, or the exercise of mental labour, skill or judgment".66

For a compilation to be protected by copyright under s 27(2) of the Copyright Act,⁶⁷ the compiler must satisfy the requirement of "intellectual creation" under s 7A(2) by exercising sufficient creativity or intellectual effort in the "selection or arrangement" of the material or data within the compilation. In respect of compilation works, it is clear

dramatic, musical, or artistic ("LDMA") work) and specific subject matter other than LDMA works: s 7. However, this will have little effect on the issues discussed here.

⁵⁹ Copyright Act (Cap 63, 2006 Rev Ed) s 27(2)(*c*).

⁶⁰ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [15].

^{61 499} US 340 (1991).

⁶² Feist Publications, Inc v Rural Telephone Service Co, Inc 499 US 340 at 352–354 (1991) ("Feist"); Jane Ginsburg, "Creation and Commercial Value: Copyright Protection of Works of Information" (1990) 90 Columbia Law Review 1865 at 1868. It was held in Feist that the defendant did not infringe any copyright by copying the plaintiff's White Pages listings because only unprotectable data was copied.

⁶³ Feist Publications, Inc v Rural Telephone Service Co, Inc 499 US 340 at 348 (1991).

^{64 [2017] 2} SLR 185.

⁶⁵ *Global Yellow Pages v Promedia Directories Pte Ltd* [2017] 2 SLR 185 at [24]: "Therefore, in the context of *compilations*, [the Court of Appeal] agree[s] ... that the 'creativity' approach is the correct one".

⁶⁶ Global Yellow Pages v Promedia Directories Pte Ltd [2017] 2 SLR 185. See also Asia Pacific Publishing Pte Ltd v Pioneers & Leaders (Publishers) Pte Ltd [2011] 4 SLR 381 at [73] (citing Telstra Corp Ltd v Phone Directories Co Pty Ltd [2010] FCA 44 at [20]) and David Tan, "Intellectual Creation in Compilations: No Sweat Required" (2018) 40 European Intellectual Property Review 338.

⁶⁷ Cap 63, 2006 Rev Ed.

No part of this document may be reproduced without permission from the copyright holders.

from ss 4, 7A and 27(2) of the current Copyright Act (read collectively) that copyright can only subsist if: (a) an author could be identified; and (b) any copyright subsisting is limited to the selection or arrangement of its contents which constitutes an intellectual creation. Sundaresh Menon CJ, delivering the unanimous judgment of the court in Global Yellow Pages, observed that "copyright protects not ideas, facts or data, but the expression thereof" [emphasis in original] and "a compilation of facts, specifically a *selection or arrangement* of the facts, was *original* and [it] could, as such, conceivably be eligible for copyright protection, though the protection conferred would be 'thin'" [emphasis in original].⁶⁸ On the facts, Menon CJ held that the arrangement of the Business Listings specifically the narrow ways in which the sorting rules departed from the default alphabetical arrangement – attracted "thin" copyright protection and would only be prima facie infringed by a near-wholesale taking of the listings.⁶⁹ This is contrasted to the finding that there was no copyright infringement concerning the taking from the Yellow Pages listings as the respondent only took "bare facts" of which copyright protection did not subsist.70

22 Generally, in the first stage of the DM process, web robots may infringe the reproduction rights of the owners in the original LDMA works if such works are copied. Copying was established in *Authors Guild v Google, Inc*,⁷¹ where books were digitised in order to make them searchable, although this was consequently held to be fair use.⁷² For instance, web robots that copy an artistic work, such as paintings, to gather information about the painting (*eg*, the number of brush strokes or the colour gradient) for further analysis, may infringe the reproduction rights to the paintings. Where web robots copy factual compilations in which the *selection* and *arrangement* of the data constitutes an intellectual creation, it would instead require a *near-wholesale* copying of that *selection* or *arrangement* before a *prima facie* case of copyright infringement can be established.

(2) Exceptions/fair dealing/fair use

There are a number of defences under the current Copyright Act that may apply to exempt liability for copyright infringement. Section 38A(1), for instance, provides that there will be no copyright

© 2021 Contributor(s) and Singapore Academy of Law.

⁶⁸ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [15].

⁶⁹ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [87].

⁷⁰ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [67].

^{71 804} F 3d 202 (2nd Cir, 2015).

⁷² *Authors Guild v Google, Inc* 804 F 3d 202 at 207 (2nd Cir, 2015), which was referred to by the Court of Appeal in *Global Yellow Pages Ltd v Promedia Directories Pte Ltd* [2017] 2 SLR 185 at [81].

No part of this document may be reproduced without permission from the copyright holders.

infringement involving a *temporary* or *transient* reproduction of a work if reproduction is made incidentally as part of the technical process of making or receiving a communication. While this appears to excuse miners from liability if the data reproduced was temporary, such an argument would unlikely succeed since Parliament had intended for the ambit of s 38A to be narrow - parties may only invoke this exception for "short-lived incidental copies" that "involve no direct action by the user" as long as there is "[no] *subsequent use* of the short-lived incidental copies" [emphasis added].⁷³ Given that DM may require the reproduction of LDMA works before using such copies to find patterns and for the purposes of predictive analytics, this reduces the prospect for miners to absolve themselves of liability with this defence. Incidentally, the fact that this defence is inapplicable when the reproduction of the work is found to be an "infringing copy of the work" under s 38A(3) puts into question whether s 38A even serves as a defence at all to copyright infringement, which may explain why there is so far no reported judgment on the reliance on this defence.

Nonetheless, miners may rely on the open-ended fair dealing defence under s 35(1) of the current Copyright Act, which recognises fair dealing for "any purpose" other than those referred to in ss 36 (for the purpose of criticism or review) and 37 (for the purpose of reporting current events). Dealings for the purpose of "research and study" may also fall within the ambit of a s 35 open-ended defence.⁷⁴ Section 35(2) then provides a non-exhaustive list of factors to be considered – and balanced – in the determination of whether a particular dealing is fair. In this regard, the Court of Appeal in *Global Yellow Pages* has provided significant guidance on the interpretation of these factors.⁷⁵ Section 35(2) in full provides:

(2) For the purposes of this Act, the matters to which regard shall be had, in determining whether a dealing with a literary, dramatic, musical or artistic work or with an adaptation of a literary, dramatic or musical work, being a dealing by way of copying the whole or a part of the work or adaptation, constitutes fair dealing with the work or adaptation for any purpose other than a purpose referred to in section 36 or 37 shall include —

(*a*) the purpose and character of the dealing, including whether such dealing is of a commercial nature or is for non-profit educational purposes;

(*b*) the nature of the work or adaptation;

⁷³ *Parliamentary Debates, Official Report* (18 July 2005), vol 80 at col 801 (S Jayakumar, Deputy Prime Minister and Minister for Law).

⁷⁴ Copyright Act (Cap 63, 2006 Rev Ed) s 35(1A).

⁷⁵ Global Yellow Pages v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [73]–[90].

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

(*c*) the amount and substantiality of the part copied taken in relation to the whole work or adaptation;

(*d*) the effect of the dealing upon the potential market for, or value of, the work or adaptation; and

(e) the possibility of obtaining the work or adaptation within a reasonable time at an ordinary commercial price.

The proposed Copyright Act 2021 contains a new s 190(1) which emphatically states that "[a] fair use of any work is a permitted use".⁷⁶ Section 191 of the Copyright Act 2021 retains the first four factors from s 35(2) of the current Copyright Act as the factors to be considered in deciding whether a work or performance is fairly used.

In *Global Yellow Pages*, the Court of Appeal traced the development of fair dealing in Singapore, as codified in the Copyright Act. The court noted that in 2004, the scope of s 35 was expanded such that a fair dealing for "any purpose" (as opposed to merely for "research or private study") might be held not to amount to an infringement of copyright, and that "[t]his also made Singapore's fair dealing provisions more similar to its American counterpart, which is more open-textured".⁷⁷ The court emphasised that the inquiry under s 35(2), "in the final analysis, is necessarily fact-sensitive".⁷⁸ Menon CJ hinted at the willingness of the local courts to take greater cognisance of American and Australian decisions in this area.⁷⁹ The persuasiveness and relevance of US fair use decisions was similarly argued in earlier academic articles.⁸⁰

27 In particular, in respect of the first factor in s 35(2), the purpose and character of the dealing, Menon CJ emphasised that "the inquiry is heavily shaped by what it was in a work that attracted copyright and what was done with that aspect of the work".⁸¹ The court referred to both English and American cases, observing that this factor favoured fair

⁷⁶ See also s 190(2)(*a*) of the Copyright Bill (Bill 17 of 2021) which provides that "[i]t is a permitted use of a protected performance to make a fair use of the performance".

⁷⁷ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [76].

⁷⁸ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [86]. Section 35(2) of the Copyright Act (Cap 63, 2006 rev Ed) enumerates five nonexhaustive fair dealing factors to be considered; four are similar to the US fair use factors, the fifth requires a consideration of "the possibility of obtaining the work or adaptation within a reasonable time at an ordinary commercial price".

⁷⁹ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [76].

⁸⁰ David Tan & Benjamin Foo, "The Unbearable Lightness of Fair Dealing: Towards an Autochthonous Approach in Singapore" (2016) 28 SAcLJ 124; David Tan, "The Transformative Use Doctrine and Fair Dealing in Singapore: Understanding the 'Purpose and Character' of Appropriation Art" (2012) 24 SAcLJ 832.

⁸¹ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [77].

dealing where "the defendant added to, recontextualised or transformed the parts taken"⁸² or where the new work was "transformative", *ie*, whether it "supersede[s] the objects" of the original creation, or "adds something new, with a further purpose or different character".83 It appears that the Court of Appeal is edging toward the view of the US Supreme Court in Campbell v Acuff-Rose Music, Inc⁸⁴ ("Campbell") when Menon CJ remarked that "we do not go as far as those cases which suggest that a commercial nature or purpose of the dealing will presumptively be regarded as unfair" and "the commerciality of the dealing is but one of the factors to be considered and it will not necessarily be fatal to a finding of fair dealing".85 In fact, the court considered the application of the transformative use doctrine in Campbell (where the commerciality of the rap song "Pretty Woman" was trumped by the transformative value of the parody) and in Authors Guild v Google, Inc⁸⁶ (where Google's making of digital copies of books for the purpose of enabling a search for identification of books containing a term of interest to the searcher

Interestingly, the court made a reference that "various circuits were apparently split"⁸⁷ on the transformative use doctrine, but unfortunately, the court did not explain further the extent to which it would accept the transformative use doctrine as informing Singapore law. The transformative use test has become the defining standard for fair use, and it has risen to the top of the agenda of the copyright academic community in the US in the last five years. In light of the US Supreme Court's decision in *Google LLC v Oracle America, Inc*⁸⁸ handed down in April 2021, the transformative use doctrine has taken a backseat in respect of the fourth factor which evaluates market impact. Justice Breyer, delivering the majority's opinion, held that: "in determining whether a use is 'transformative', we must go further and examine the copying's more specifically described 'purpose[s]' and 'character^{".89} Furthermore, the court would "take into account the public benefits the copying will likely

involved a highly transformative purpose).

88 141 S Ct 1183 (2021).

⁸² Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [79] (where the Court of Appeal referred to Newspaper Licensing Agency v Marks & Spencer plc [2003] 1 AC 551; [1999] EMLR 369 at 380 and University of London Press Ltd v University Tutorial Press Ltd [1916] 2 Ch 601 at 613–614).

⁸³ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [79] (where the Court of Appeal referred to Campbell v Acuff-Rose Music, Inc 510 US 569 (1994)).

⁸⁴ Campbell v Acuff-Rose Music, Inc 510 US 569 (1994).

⁸⁵ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [81].

⁸⁶ Authors Guild v Google, Inc 804 F 3d 202 (2nd Cir, 2015).

⁸⁷ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [88].

⁸⁹ *Google LLC v Oracle America, Inc* 141 S Ct 1183 at 1203 (2021).

^{© 2021} Contributor(s) and Singapore Academy of Law.

produce".⁹⁰ The majority concluded that since Google had reimplemented a user interface, taking only what was needed to allow users to put their accrued talents to work in a new and transformative program, its copying of the Sun Java API was a fair use of the material.

29 The second factor in s 35(2), the nature of the work, recognises that some works are closer to the core of intended copyright protection than others, especially LDMA works that are highly creative in nature.⁹¹ This factor tends to weigh against fair use when LDMA works are being copied. As for the third factor in s 35(2), the amount and substantiality of the portion used in relation to the copyrighted work, it asks whether the quantity and value of the materials used are reasonable in relation to the purpose of the copying.⁹² Nonetheless, there is no blanket rule against copying entire works where such copying is *reasonably necessary*.⁹³

30 The third factor also influences the fourth factor in s 35(2), which investigates the effect of the use of original work on the potential market on the value of the copyrighted work. This requires the court to consider "not only the extent of market harm caused by" the alleged infringer's action, but also whether the defendant's conduct, if "unrestricted and widespread", would "result in a substantially adverse impact on the potential market" for the original; it takes into account not only harm to the original but also harm to the market for derivatives works.⁹⁴ The fourth factor generally seeks to uphold the incentive rationale that underpins copyright law by preventing unjust enrichment and harm to the original works, thereby "facilitat[ing] greater investment, research and development in copyright industries in Singapore".⁹⁵ This fourth

⁹⁰ *Google LLC v Oracle America, Inc* 141 S Ct 1183 at 1206 (2021).

⁹¹ Staniforth Ricketson & Christopher Creswell, *The Law of Intellectual Property: Copyright, Designs and Confidential Information* (Lawbook Co, 2nd Ed, 2002). See also *Campbell v Acuff-Rose Music, Inc* 510 US 569 at 586 (1994): "This factor calls for recognition that some works are closer to the core of intended copyright protection than others, with the consequence that fair use is more difficult to establish when the former works are copied."

⁹² Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [83].

⁹³ For the US position, see Cariou v Prince 714 F 3d 694 at 710 (2nd Cir, 2013) (which cited Bill Graham Archives v Dorling Kindersley Ltd 448 F 3d 605 at 613 (2nd Cir, 2006)):

Although neither our court nor any of our sister circuits has ever ruled that the copying of an entire work favours fair use, ... courts have concluded that such copying does not necessarily weigh against fair use because *copying the entirety of a work is sometimes necessary* to make a fair use of the image. The third-factor inquiry must take into account that the *extent of permissible copying varies with the purpose and character of the use.* [emphasis added]

⁹⁴ Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [84].

⁹⁵ Parliamentary Debates, Official Report (16 November 2004), vol 78 at col 1052 (S Jayakumar, Deputy Prime Minister and Minister for Law): Parliament (cont'd on the next page)

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

factor is influenced by the degree of transformative use present under the first factor; a finding of a highly transformative use will result likely in a finding of little or no market substitution, and market harm will not be so readily inferred. However, where there is moderate or little transformative use, the greater the quantity or quality taken (third factor) may indicate that the secondary work serves as a market substitute, with the fourth factor thus weighing against fair use. The US Second Circuit Court of Appeals in TCA Corp v McCollum⁹⁶ has commented that the generous view of what might constitute transformative use (and therefore fair use) might have hit its "high-water mark" in Cariou v Prince,97 and Judge Pierre Leval, now sitting on the US Second Circuit Court of Appeals, has retreated noticeably from endorsing the transformative use talisman in Capitol Records, LLC v ReDigi Inc⁹⁸ commenting that the fourth factor is "undoubtedly the single most important element of fair use".⁹⁹ Judge Leval correctly observed that the fourth factor is a consideration of whether the secondary use brings a competing substitute to the marketplace, and "the more the objective of the secondary use differs from the original, the less likely it will be to supplant the commercial market for the original" [references omitted].¹⁰⁰ The renaissance of the primacy of the fourth factor was alluded to in the majority's judgment of the US Supreme Court in Google LLC v Oracle America, Inc,¹⁰¹ and more forcefully emphasised in the dissenting judgment.¹⁰²

31 Finally, the fourth factor possibly influences the fifth factor in s 35(2), which evaluates the possibility of obtaining the work within a reasonable time at an ordinary commercial price.¹⁰³ Essentially, if the defendant could have obtained the work on reasonable commercial terms, then this factor weighs against fair use. The fifth factor, however,

commissioned a study which considered jurisdictions including the UK, Australia, Canada, Germany, France and the US.

- 97 714 F 3d 694 (2nd Cir, 2013).
- 98 910 F 3d 649 (2nd Cir, 2018).

100 Capitol Records, LLC v ReDigi Inc 910 F 3d 649 at 662 (2nd Cir, 2018).

⁹⁶ TCA Corp v McCollum 839 F 3d 168 at 181 (2nd Cir, 2016). See also Kienitz v Sconnie Nation, LLC 766 F 3d 756 at 758 (7th Cir, 2014), per Judge Easterbrook of the US Seventh Circuit Court of Appeals, who was also highly critical of the Cariou decision: "[w]e're skeptical of Cariou's approach, because asking exclusively whether something is 'transformative' not only replaces the list in § 107 but also could override 17 USC § 106(2), which protects derivative works".

⁹⁹ Capitol Records, LLC v ReDigi Inc 910 F 3d 649 at 662 (2nd Cir, 2018) (citing Harper & Row Publishers, Inc v Nation Enterprises 471 US 539 at 566 (1985)).

¹⁰¹ Google LLC v Oracle America, Inc 141 S Ct 1183 at 1206–1209 (2021).

¹⁰² Google LLC v Oracle America, Inc 141 S Ct 1183 at 1216.

¹⁰³ See generally, Global Yellow Pages Ltd v Promedia Directories Pte Ltd [2017] 2 SLR 185 at [35] and David Tan & Benjamin Foo, "The Unbearable Lightness of Fair Dealing: Towards an Autochthonous Approach in Singapore" (2016) 28 SAcLJ 124 at paras 44–47.

No part of this document may be reproduced without permission from the copyright holders.

will no longer be relevant as Proposal 6 of the Copyright Review Report has recommended the removal of the final factor, in order to "mirror more closely in form"¹⁰⁴ the US fair use provision that only consists of the first four factors. This has also been affirmed in s 191 of the proposed Copyright Act 2021 where this factor has been removed.

(3) Fair use and data mining – The American experience

32 A number of decisions of the US Circuit Courts of Appeals are apropos in providing guidance on how DM in Singapore may be treated under a general fair use provision. Regarding the first factor, HathiTrust¹⁰⁵ is instructive - the issue was whether the digitisation of copyrighted works by 13 universities and other organisations in creating the HathiTrust Digital Library ("HDL") without authorisation may constitute fair use. The Second Circuit Court of Appeals found that the first factor weighed in favour of fair use as HDL's enabling of full-text search "serves a new and different function from the original" and contributes to the public benefit.¹⁰⁶ Additionally, the dealing was found to carry a "non-profit educational" purpose as the HDL was a project started by educational and non-profit institutions targeted at providing greater access to works without any "purely commercial" motive.¹⁰⁷ Even if there is a commercial motivation, the Second Circuit Court of Appeals in Authors Guild v Google, Inc saw "no reason ... why Google's overall profit motivation should prevail as a reason for denying fair use over its highly convincing transformative purpose, together with the absence of significant substitutive competition, as reasons for granting fair use".¹⁰⁸ The court held that similar to HathiTrust, the purpose of Google's copying of the original copyrighted books is "to make available significant information about those books, permitting a searcher to identify those that contain a word or term of interest, as well as those that do not include reference to it"¹⁰⁹ [emphasis in original] which is significantly different from the purposes of the original books.

33 These decisions lend clear support that the first factor would apply favourably to DM activities in favour of fair use in Singapore, due

¹⁰⁴ Ministry of Law and Intellectual Property Office of Singapore, *Singapore Copyright Review Report* (17 January 2019) at para 2.6.8.

¹⁰⁵ Authors Guild, Inc v HathiTrust 755 F 3d 87 at 92 (2nd Cir, 2014).

¹⁰⁶ Authors Guild, Inc v HathiTrust 755 F 3d 87 at 97 (2nd Cir, 2014) ("HathiTrust"). See also William F Patry, *Patry on Copyright* vol 4 (West, Online, 2015) at \$10:21 (observing that the use in *HathiTrust* is "socially beneficial, serves a different purpose than the original, and is in no way substitutional").

¹⁰⁷ Authors Guild, Inc v HathiTrust 755 F 3d 87 at 90–91 (2nd Cir, 2014).

¹⁰⁸ Authors Guild v Google, Inc 804 F 3d 202 at 219 (2nd Cir, 2015).

¹⁰⁹ Authors Guild v Google, Inc 804 F 3d 202 at 217 (2nd Cir, 2015).

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

to DM's clear public benefit of identifying patterns from raw works from which new knowledge can be derived. Although the *commerciality* of the DM activity itself – whether DM is used for a non-profit or a commercial purpose – is a relevant consideration in the first factor analysis, the US decisions suggest that it is not important since many of the most universally accepted forms of fair use, such as news reporting and commentary, reviews of books, and performances, as well as parody, are all normally done commercially for profit.¹¹⁰

The second factor would likely weigh against finding fair use in DM, but it is unlikely to have significant weight in the overall analysis. In *HathiTrust*, the court rejected the argument that the second factor should weigh against fair use given that the millions of works digitised would almost certainly contain works of creative endeavour, as opposed to factual compilations.¹¹¹ Nevertheless, the Second Circuit Court of Appeals in *Authors Guild v Google, Inc*,¹¹² commented that it "has rarely played a significant role in the determination of a fair use dispute".¹¹³

The Ninth Circuit Court of Appeals decision in Kelly v Arriba 35 Soft Corp¹¹⁴ is also useful in understanding how the evaluation of the third factor could be applied to DM. There, it was held that the use of entire copyrighted works was *necessary* in cases involving search engines since copying only a part of the copyrighted work would create practical difficulties for users, thereby reducing the usefulness of the search engine. In the same vein, even if entire works were copied by web robots in the DM context, it could be reasoned that such a taking is reasonable, considering the *different* purpose of the dealing (*ie*, to identify patterns in vast amounts of raw data); thus, the third factor might not necessarily weigh against fair use. The Second Circuit Court of Appeals has observed that the courts have rejected any categorical rule that a copying of the entirety cannot be fair use, especially when the copying was reasonably appropriate to achieve the copier's transformative purpose and was conducted in a manner that did not offer a competing substitute for the original.¹¹⁵ It is important that in DM activities, the entire original work is not revealed to the public but retained by the miner to enable

¹¹⁰ Authors Guild v Google, Inc 804 F 3d 202 at 219 (2nd Cir, 2015).

¹¹¹ David Tan & Benjamin Foo, "The Unbearable Lightness of Fair Dealing: Towards an Autochthonous Approach in Singapore" (2016) 28 SAcLJ 124 at para 53.

^{112 804} F 3d 202 (2nd Cir, 2015).

¹¹³ Authors Guild v Google, Inc 804 F 3d 202 at 219 (2nd Cir, 2015).

^{114 336} F 3d 811 (9th Cir, 2003).

¹¹⁵ Authors Guild v Google, Inc 804 F 3d 202 at 220 (2nd Cir, 2015).

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

the particular analysis to be undertaken. This would be analogous to the Google Books scenario.¹¹⁶

36 The application of the fourth factor to DM is highly dependent on the finding of the first factor. The US Supreme Court in Campbell had emphasised the close linkage between the first and fourth factors, in that the more the copying is done to achieve a purpose that differs from the purpose of the original, the less likely it is that the copy will serve as a satisfactory substitute for the original.¹¹⁷ The Second Circuit Court of Appeals noted that even if the *purpose* of the copying was for a valuably transformative purpose, such copying might nonetheless harm the value of the copyrighted original if done in a manner that resulted in widespread revelation of sufficiently significant portions of the original as to make available a significantly competing substitute.¹¹⁸ Generally, copyrighted works copied for DM purposes will require extensive processing and analysis before knowledge is derived and shared. Miners must ensure that they do not reveal significant portions of the original copyrighted works to the public. Although one could argue that DM could limit the rights owners' expansion into a potential market (eg, a lost opportunity to license the works¹¹⁹) since markets are dynamic and change over time to meet new demands, the US Circuit Courts have universally dismissed this argument where only a small portion of the original works was revealed to the public. In Authors Guild v Google, Inc,¹²⁰ the Second Circuit Court of Appeals held that "a mere revelation of 16% of the text of plaintiffs' books overstates the degree to which snippet view can provide a meaningful substitute".121

37 The more recent decision of the Ninth Circuit Court of Appeals in *VHT*, *Inc v Zillow Group*, *Inc*¹²² in 2019 had the opportunity to discuss how DM activities may be permitted under a fair use analysis. The court considered how the Google Books search engine enables a full-text search, which allows users to search for a specific term, and then provides "snippets", or a part of a page, for users to read. The highly transformative universal search function was seen to augment public knowledge by

¹¹⁶ Authors Guild v Google, Inc 804 F 3d 202 at 221–222 (2nd Cir, 2015): "While Google makes an unauthorized digital copy of the entire book, it does not reveal that digital copy to the public. The copy is made to enable the search functions to reveal limited, important information about the books" [emphasis in original].

¹¹⁷ Campbell v Acuff-Rose Music, Inc 510 US 569 at 591 (1994).

¹¹⁸ Authors Guild v Google, Inc 804 F 3d 202 at 223 (2nd Cir, 2015).

¹¹⁹ Authors Guild, Inc v HathiTrust 755 F 3d 87 at 99 (2nd Cir, 2014) (this was an argument the plaintiffs raised).

^{120 804} F 3d 202 (2nd Cir, 2015).

¹²¹ Authors Guild v Google, Inc 804 F 3d 202 at 223 (2nd Cir, 2015).

^{122 918} F 3d 723 (9th Cir, 2019).

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

making available information *about* books without providing the public with a substantial substitute. Furthermore, the search engine was perceived to be making possible "a new type of *research* known as 'text mining' or 'data mining', whereby users can search across the corpus of books to determine the frequency of specified terms across time".¹²³ Indeed DM can be used for a kaleidoscope of purposes different from the original works, and the fair use factors are sufficiently comprehensive in their scope to be effective in evaluating any new scenario.

38 Given how these four open-ended fair use/fair dealing factors can be applied to the DM context with ease, this begs the question of whether a new specific DM exception in s 244 of the proposed Copyright Act 2021 laden with multiple conditions to be fulfilled is even necessary.

(4) Issues with the new computational data analysis exception

In implementing Proposal 8 of the Copyright Review Report, s 244 of the proposed Copyright Act 2021 introduces a specific exception for reproduction of works made for the purpose of *computational data analysis* provided that the five conditions in ss 244(2)(a) to 244(2)(e) are met. Computational data analysis is defined non-exhaustively as "using a computer program to identify, extract and analyse information or data from the work" – which is synonymous with DM.¹²⁴ In particular, the sole requirement under Proposal 8 has also been transposed to s 244, which requires that there must be "lawful access" to the works copied before the specific exception can apply. Currently, the two illustrations in s 244 of the proposed Copyright Act 2021 give us a limited understanding of what "lawful access" entails – they classify the act of bypassing TOS agreements and circumventing paywalls of a database as unlawful.

40 However, these illustrations are inadequate and there is significant ambiguity of what "lawful access" means in a number of situations. For instance, it is unclear whether bypassing REPs would constitute *unlawful* access by the miner, especially since REPs are frequently used by owners to protect their works and are easily adoptable from online repositories such as GitHub.¹²⁵ Potentially beneficial uses of DM may also not be a permitted use under this new exception. One notable example includes useful applications of DM such as smart disclosure systems ("SDSs") which allow users to have timely access to pre-contractual information to

© 2021 Contributor(s) and Singapore Academy of Law.

¹²³ VHT, Inc v Zillow Group, Inc 918 F 3d 723 at 742 (9th Cir, 2019).

¹²⁴ Copyright Bill (Bill 17 of 2021) s 243(*a*).

¹²⁵ See for instance BrandwatchLtd's Github page https://github.com/BrandwatchLtd/ robots> (accessed 17 May 2021).

No part of this document may be reproduced without permission from the copyright holders.

better understand the terms of the user agreement.¹²⁶ Put simply, if a new "bionic eye" is available to scan a contractual document and extract the relevant information for users to understand the terms and conditions relating to the use of the website, it would arguably be unjustified to prohibit such use.¹²⁷ Introducing the specific exception in the Copyright Act, *without more*, may hence even do more harm than good in the foreseeable future. It may also unduly prejudice a miner who failed to qualify for protection under this computational data analysis exception when the miner attempts to advance an alternative argument under the open-ended fair use provision in ss 190 to 191 of the proposed Copyright Act 2021. Accordingly, a re-examination of this specific exception is both timely and necessary.

B. Data protection laws

(1) Obligations under Personal Data Protection Act

41 Apart from creative authorial works which do not contain personal data, the growing popularity of DM can be a threat to the security of an individual's sensitive information which are contained in compilations. In turn, from a strict legal perspective, the PDPA's role in regulating DM cannot be understated, since DM may also attract liability under the PDPA when it involves the unlawful collection, use, and/or disclosure of personal data – data that can identify an individual (or data subject) on its own, or with other information to which the organisation has or is likely to have access.¹²⁸ Essentially, the relationship between the Copyright Act and the PDPA is a largely symbiotic one demarcating opposite ends of the fact-expression dichotomy; while the Copyright Act protects expressions that constitute an intellectual creation, the PDPA protects facts that constitute personal data.

No part of this document may be reproduced without permission from the copyright holders.

¹²⁶ Essentially, one of the main goals of smart disclosure systems is to increase the awareness of users of the rights, obligations and possible risks in their online activities, and to mitigate the consequences of the well-known signing-without-reading process (which happens especially often with click-wrap agreements).

¹²⁷ Rossana Ducato & Alain M Strowel, "Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to Machine Legibility" (2019) 50 International Review of Intellectual Property & Competition Law 649.

¹²⁸ Personal Data Protection Act 2012 (Act 26 of 2012) s 2. See also Personal Data Protection Commission, *Advisory Guidelines on Key Concepts in the Personal Data Protection Act* (23 September 2013) at para 5.4, which states that personal data includes information about an individual's health, educational and employment background, as well as an individual's activities such as spending patterns.

⁴²Broadly speaking, the PDPA contains 11 main obligations that organisations¹²⁹ are required to comply if they undertake activities relating to the collection, use or disclosure of personal data.¹³⁰ Should an organisation fail to discharge any of these obligations, s 48O of the newly amended PDPA provides the grounds on which a private claim can be brought against offenders responsible for such misconduct. In a DM context, certain contentious obligations are hence likely to be in issue, such as: (a) whether notification was given to the individual to inform them of the purposes of the collection, use, or disclosure of one's personal data ("notification obligation");¹³¹ (b) whether consent was given by the individual for one's personal data to be collected, used or disclosed ("consent obligation");¹³² and (c) the organisation is collecting, using or disclosing personal data for purposes that a reasonable person would consider "appropriate" ("purpose limitation obligation").¹³³

(2) Exceptions under Personal Data Protection Act in a DM context

43 In the absence of guidance from case law, it is unclear how these PDPA obligations could potentially impact DM activities. Firstly, miners may find it impossible to satisfy the notification obligation, which requires them to inform the individual the purposes for which one's personal data will be collected, used or disclosed before the actual collection, use or disclosure.¹³⁴ This is because it may not be possible to contact some individuals, especially if the personal data was collected from a large sample size. Even unique identifiers that can identify individuals, such as one's full name, do not provide adequate means to contact them.¹³⁵ In the unlikely scenario that this gargantuan task of contacting all identifiable individuals was surmountable, "notification" may still be inadequate since miners do not (and cannot) know in advance what they may discover from the DM process.

¹²⁹ Although the Personal Data Protection Act 2012 (Act 26 of 2012) uses the word "organisation", "organisation" includes any individual, company, association or body of persons: s 2. As such, the use of "organisations" in this segment of the paper will be used interchangeably with miners who will also have to comply with the same obligations.

¹³⁰ These obligations include the consent obligation, the purpose limitation obligation, the notification obligation, the access and correction obligation, the accuracy obligation, the protection obligation, the retention limitation obligation, the transfer limitation obligation, the accountability obligation, the data breach notification obligation and the data portability obligation.

¹³¹ Personal Data Protection Act 2012 (Act 26 of 2012) s 20.

¹³² Personal Data Protection Act 2012 (Act 26 of 2012) ss 13 and 14.

¹³³ Personal Data Protection Act 2012 (Act 26 of 2012) s 18(*a*).

¹³⁴ Personal Data Protection Act 2012 (Act 26 of 2012) s 20(1).

¹³⁵ Personal Data Protection Commission, *Advisory Guidelines on Key Concepts in the Personal Data Protection Act* (23 September 2013) at para 5.10.

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

Secondly, it is also onerous to discharge the consent obligation, since it resembles an opt-in regime where the individual has to "opt-in" to consent to one's personal data being collected, used or disclosed.¹³⁶ Failure to fulfil the notification obligation also impacts the discharge of the consent obligation, since s 14(1)(a) requires the data subject to be *notified* of the purposes for which his or her personal data is collected, used or disclosed. This means that the consent obligation also suffers from the same limitations as the notification requirement: that it may not be practicable to contact every identifiable individual at every stage of the DM process. These difficulties also negate the possibility of proving deemed consent by notification under the newly introduced s 15A, which also requires notifying these individuals of the purpose of the organisation's intention to collect, use or disclose their personal data.¹³⁷

45 Finally, to complicate matters, even assuming that the consent obligation was discharged, individuals are still free to withdraw their consent at any stage of the DM process, be it either during the collection, use or disclosure of their personal data.¹³⁸ These implications are therefore tremendous in the DM context – this could effectively restrict *most*, if not *all*, DM activities that involves the collection, use and disclosure of personal data.

Granted, the notification and consent obligations can be negated by proving that the personal data is *publicly available*.¹³⁹ Section 2(1) defines this to mean personal data about an individual that is generally available to the public, including personal data which can be observed by *reasonably expected means* at a location or an event *open to the public* at which the individual appears. For example, disclosure of personal data to a closed online group where members of the public could join with *minimal* effort may amount to making the personal data publicly available.¹⁴⁰ It follows that for most DM activities that involve datasets on webpages that require *minimal* or *no effort* to access, the "publicly available" exception may apply to negate the consent obligation.

47 However, the same difficulty regarding the specific computational data analysis exception under the Copyright Act rears its ugly head once again under the PDPA: should the publicly available exception be available as a defence when owners have already reserved their rights to the contrary (*eg*, by way of TOS agreements or REPs)? Once again, we

© 2021 Contributor(s) and Singapore Academy of Law.

¹³⁶ Personal Data Protection Act 2012 (Act 26 of 2012) s 14(1).

¹³⁷ Personal Data Protection Act 2012 (Act 26 of 2012) s 15A(4)(*b*)(i).

¹³⁸ Personal Data Protection Act 2012 (Act 26 of 2012) s 16.

¹³⁹ Personal Data Protection Act 2012 (Act 26 of 2012) First Schedule, Pt 2, para 1.

¹⁴⁰ Personal Data Protection Commission, *Advisory Guidelines on Key Concepts in the Personal Data Protection Act* (23 September 2013) at para 12.59.

No part of this document may be reproduced without permission from the copyright holders.

find ourselves in uncharted waters. While the Personal Data Protection Commission had proposed a multi-factorial approach to consider the ease with which the public can gain access to the place (*ie*, the presence of physical barriers, the conditions and effectiveness of these barriers, the employment of security systems, and sentries and patrols aimed at restricting entry¹⁴¹), these factors were clearly intended to apply in the context of physical spaces, instead of digital platforms like webpages.

48 Even though it can be argued that similar factors can be adopted to evaluate the accessibility of raw data on webpages, this would require an in-depth analysis of whether TOS agreements or REPs serve as *effective* digital barriers to prevent the exploitation of web domains by web robots. Much would depend on the conspicuousness of the TOS agreements on the webpage, or the coding prowess of REPs to weed out web robots, and such considerations are necessarily fact-centric.

Unfortunately, the recourses available - to negate the consent 49 and notification obligations - are lines of reasoning yet to be explored in Singapore case law. The First Schedule under the PDPA now allows an organisation to collect, use and disclose personal data about an individual without consent (and without having to notify) where such activities are in the *legitimate interests* of the organisation, and the legitimate interests outweigh any adverse effect on the individual.¹⁴² Additionally, while there are specific types of legitimate interests enumerated in the First Schedule, those that apply to the DM context are *narrow* in scope: (a) the provision of legal services by the organisation to another individual;¹⁴³ (b) allowing the organisation to obtain legal services;¹⁴⁴ or (c) enabling the organisation to provide a service for the personal or domestic purposes for a separate individual who had provided the personal data of the identifiable individual.¹⁴⁵ Nonetheless, it is not often that the adverse effect on the individual for the collection, use or disclosure of their personal data is minimal (especially sensitive personal data), and neither is it commonplace for DM activities to fit the bill for any of the specific scenarios. In this light, its utility for miners is likely limited in scope should they wish to negate the application of the consent and notification obligations.

50 The better recourse may be to rely on the research exception, which aims to "support commercial research and development that is

No part of this document may be reproduced without permission from the copyright holders.

¹⁴¹ Personal Data Protection Commission, *Advisory Guidelines on Key Concepts in the Personal Data Protection Act* (23 September 2013) at para 12.64.

¹⁴² Personal Data Protection Act 2012 (Act 26 of 2012) First Schedule, Pt 3, para 1(1)(b).

¹⁴³ Personal Data Protection Act 2012 (Act 26 of 2012) First Schedule, Pt 3, para 5.

¹⁴⁴ Personal Data Protection Act 2012 (Act 26 of 2012) First Schedule, Pt 3, para 5.

¹⁴⁵ Personal Data Protection Act 2012 (Act 26 of 2012) First Schedule, Pt 3, para 8.

^{© 2021} Contributor(s) and Singapore Academy of Law.

not immediately directed at productisation".¹⁴⁶ Under this exception, the consent and notification obligations for the *use* of personal data can be negated, provided that:¹⁴⁷

(*a*) the research purpose cannot reasonably be accomplished unless the personal data is used in an individually identifiable form;

(b) there is a clear public benefit to using the personal data for the research purpose;

(*c*) the results of the research will not be used to make any decision that affects the individual; and

(d) in the event that the results of the research are published, the organisation should publish the results in a form that does not identify the individual.

⁵¹ The research exception likewise applies for *disclosure* of personal data, lest the additional requirement mandating that it is "impracticable for the organisation to seek the consent of the individual [during the disclosure process]".¹⁴⁸ Ideally, this applies to "research institutes carrying out scientific research and development, educational institutes embarking on social sciences research, and organisations conducting market research to identify and understand potential customer segments", although this list is non-exhaustive.¹⁴⁹ Understandably, this expansive exception could encompass the bulk of DM activities dealing with personal data.

52 Yet, the shift towards softer regulation is no cause for unfettered celebration. Despite greater latitude being given to DM activities under these amendments, two avenues are still worth reviewing. First, the amendments reaffirm the status quo (before the 2020 amendments) that the research exception would only apply for the *use* and *disclosure* of personal data in DM, but not its *collection*. This would mean that

¹⁴⁶ Parliamentary Debates, Official Report (2 November 2020), vol 95 (S Iswaran, Minister for Communications and Information and Minister-in-charge of Trade Relations).

¹⁴⁷ Personal Data Protection Act 2012 (Act 26 of 2012) ("PDPA") Second Schedule, Pt 2, Div 3. Although not expressly stated in the Second Schedule, the research exception can also negate the notification obligation. Section 20(3)(b) of the PDPA states that the notification obligation does not apply if the organisation fulfils the requirements under s 17. Section 17 specifies the instances where the organisation can collect, use, or disclose the personal data *without the consent of the individual* but nonetheless in accordance with Pt 2 of the Second Schedule (s 17(1)(b)) or Pt 3 of the Second Schedule (s 17(1)(c)). The research exception, which can be found under both Pts 2 and 3 of the Second Schedule, would hence not only negate the consent obligation, but also the notification obligation (s 20(3)(b) read with s 17).

¹⁴⁸ Personal Data Protection Act 2012 (Act 26 of 2012) Second Schedule, Pt 3, Div 2, para 1(*b*).

¹⁴⁹ Parliamentary Debates, Official Report (2 November 2020), vol 95 (S Iswaran, Minister for Communications and Information and Minister-in-charge of Trade Relations).

No part of this document may be reproduced without permission from the copyright holders.

miners must still discharge the consent and notification obligations when *collecting* personal data, which still runs into the problems of contactability and inadequate notification, as explained earlier.¹⁵⁰ It follows that if the consent or notification obligations cannot be waived at the *collection* stage, liberalising the subsequent *use* and *disclosure* of personal data would be nothing more than a paper tiger, since DM activities would not be lawful from the outset during the collection process. Moreover, it remains unclear whether miners can rely on the research exception despite reservation rights to the contrary using common industry practices like TOS agreements and REPs. With this cloud of uncertainty left hanging, miners may be deterred from engaging in related DM activities without assurance that the research exception can apply in their favour.

C. Interim observations – Potential reforms

As it stands, the Copyright Act remains the primary regulator for DM, except when such activities concern the collection, use or disclosure of personal data, which would instead be governed by the PDPA. The table below presents the scope of the Copyright Act and PDPA in regulating DM activities, and *how* provisions within these legal regimes might be applicable.

| Statute | What is Protected | Legal Principles |
|---------------------------------------|---|--|
| Copyright Act | expressions with sufficient creative input ("works") the <i>selection</i> and <i>arrangement</i> of data in factual compilations | Mandates that copyright (<i>eg, reproduction rights</i>) should not be infringed. Open-ended fair dealing/fair use defence or new specific exception may apply. |
| Personal Data Protection Act | data that constitutes <i>personal data</i> | Mandates that collection, use or disclosure of personal data must satisfy "notification" and "consent" obligations. "Legitimate interest" or broader "research" exception may apply to negate the consent and notification obligations. |

54 It is nonetheless clear that a plethora of legal issues is left unanswered. Firstly, the definition of "lawful access" under the specific

¹⁵⁰ See paras 43-44 above.

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

exception¹⁵¹ remains a confusing terminology in certain situations such as when REPs are adopted by owners. Secondly, it potentially excludes legitimate uses of DM that taps on useful applications, like SDSs that allow miners to better understand the terms of use of a webpage. One could argue that an introduction of such an exception is short-sighted, and instead signifies an attempt to provide an *illusion of certainty* in a situation that legislators may not completely understand.

Likewise, even though the scope of the research exception may be wide enough to encompass the kaleidoscope of DM activities, this is arguably inconsequential since it does not waive the consent or notification obligations for miners during the *collection* of personal data. Finally, whether the research exception should be subject to reservations by owners to the contrary, such as through TOS agreements or REPs, remains up for debate and requires further illumination.

V. Stakeholders' perspectives

A. General preference for the open-ended fair dealing/fair use defence over computational data analysis exception

⁵⁶ The findings from the interviews conducted indicated that all 11 Category A respondents¹⁵² and six out of 11 Category B respondents¹⁵³ favour the open-ended fair dealing defence in the DM context; this seems contrary to what the Copyright Review Report suggests.¹⁵⁴ Across both categories of respondents, their description of the "lawful access" requirement under the proposed DM exception ranged from "narrow"¹⁵⁵ to "vague"¹⁵⁶ and "fuzzy".¹⁵⁷ One content creator noted that this requirement failed to consider situations where it may not be within the owner's intention to allow mining of their online content even though lawful access was given for the purpose of pure consumption and enjoyment of the content.¹⁵⁸ Unsurprisingly, a miner questioned whether the use

¹⁵¹ This refers to both Proposal 8 of the Copyright Review Report and s 244 of the Copyright Bill (Bill 17 of 2021) which are similar since they retain the "lawful access" requirement.

¹⁵² Interview with all Category A respondents from A1 to A11.

¹⁵³ Interview with B2, B4, B5, B6, B7 and B9.

¹⁵⁴ Ministry of Law and Intellectual Property Office of Singapore, *Singapore Copyright Review Report* (17 January 2019) at para 2.8.1: it was indicated that there was "majority support" for introducing the new DM exception.

¹⁵⁵ Interview with B8.

¹⁵⁶ Interview with B3.

¹⁵⁷ Interview with A2.

¹⁵⁸ Interview with B4.

 $[\]ensuremath{^{\odot}}$ 2021 Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

of REPs could potentially affect the finding of "lawful access".¹⁵⁹ These responses highlight how DM stakeholders in Singapore have significant concerns about the ambiguity of the "lawful access" requirement.

⁵⁷Of all those who favour the open-ended fair dealing/fair use defence, two Category A respondents were of the view that the proposed requirement of "lawful access" *overly assumes* that stakeholder dynamics in a DM context is always of a commercial nature, but fails to consider non-commercial relationships where owners do not actively prescribe ways to access their work through commercial means.¹⁶⁰ Where access to such online content seems freely available and absent of any firewall, these respondents worry that the specific DM exception would unnecessarily encourage the mining of such works.¹⁶¹

58 Regarding TOS agreements, three Category B respondents¹⁶² noted that there are multiple ways to display such agreements, ranging from "pop-ups"¹⁶³ to "a separate tab"¹⁶⁴ where the terms of use can be viewed. One such respondent suggested that the requirement of "lawful access" might involve a fact-specific exercise that requires the court to adopt a "reasonableness" test to determine whether a miner reasonably conducting due diligence of the webpage would likely have been privy to such terms of use.¹⁶⁵

59 More broadly, six respondents across both categories supported the open-ended fair dealing/fair use defence because it appeared rooted in the principle of utilitarianism, given that the first fair use factor posed a pertinent inquiry as to whether the alleged infringing work was for a different purpose or was transformative.¹⁶⁶ One miner, in particular, aptly analogised the strict requirements of the specific DM exception with an example in patent law – that if contact tracing in a COVID-19 pandemic was patentable by a private company in Singapore, this will result in public harm on a global scale if the only applicable defence against patent infringement requires other countries to get "lawful access" to the patent.¹⁶⁷ Even the five Category B respondents who preferred the

- 162 Interview with B5, B7 and B9.
- 163 Interview with B5, B7 and B9.
- 164 Interview with B5 and B7.
- 165 Interview with B5.
- 166 Interview with A6, A5, A7, A9, B2 and B4.
- 167 Interview with A6.

¹⁵⁹ Interview with A8.

¹⁶⁰ Interview with A3 and A5.

¹⁶¹ Interview with A3 and A5.

specific DM exception nonetheless acknowledged that it may negatively impact the overall public benefit.¹⁶⁸

Nonetheless, the same five Category B respondents who prefer the specific exception highlighted that this specific exception would be beneficial as it can function as a *stronger* deterrent in situations where it can be established that the mining was *clearly* unlawful.¹⁶⁹ These usually occur in cases involving outright theft of works when alternative methods of procuring these works (*eg*, an option to purchase or obtain a license) are readily available.¹⁷⁰ While the same outcome may be reached under the open-ended fair dealing exception by a finding that the fourth factor weighed against fair dealing, three of such respondents pointed out that unlawful DM activities (ousting the applicability of the specific exception) may still be considered fair dealing under the open-ended defence if the work was still *highly* transformative enough under the first factor.¹⁷¹

61 However, on the issue of whether the specific exception would potentially restrict other legitimate uses of DM, no respondent from either category was able to think of such examples, besides conventional methods of mining. In fact, none of the respondents have ever encountered SDSs.¹⁷² While our findings are not a representative sample of the entire DM community, it at least suggests that this legal uncertainty should not be a prominent factor influencing the need for reform in this area.

B. Preference to preserve status quo for the research exception in Personal Data Protection Act

62 One significant theme that emerged from *all* Category A and B respondents is a sense of unease at the thought of having their personal data used *completely* without consent.¹⁷³ Some respondents described such situations as "creepy"¹⁷⁴ and "disturbing".¹⁷⁵ Specifically, 13 out of 22 respondents justified the consent requirement on grounds that misuse of personal data could lead to irreversible harms for the identifiable individual.¹⁷⁶ In line with their endorsement that consent is necessary, all 22 respondents were in consensus for preserving the status quo for the research exception of the PDPA.

¹⁶⁸ Interview with B1, B3, B7, B8 and B10.

¹⁶⁹ Interview with B1, B3, B7, B8 and B10.

¹⁷⁰ Interview with B7.

¹⁷¹ Interview with B1, B3 and B8.

¹⁷² Interview with all Category A and B respondents including A1 to A11 and B1 to B11.

¹⁷³ Interview with all Category A and B respondents including A1 to A11 and B1 to B11.

¹⁷⁴ Interview with B4.

¹⁷⁵ Interview with A3.

¹⁷⁶ Interview with A2, A3, A6, A7, A8, A9, A10, B1, B3, B4, B7, B8 and B10.

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

⁶³ Two respondents further justified their views on grounds that the research exception is seemingly broad, since many DM activities can be classified as "research".¹⁷⁷ Without the consent requirement, it potentially leads to abuse of the research exception such that organisations would collect and use personal data for DM purposes under the guise of "research" when it was in fact purely for their own commercial gain.¹⁷⁸

All 22 respondents also agreed that the research exception should be subject to reservations by the data owner to the contrary, especially in cases where they are easily accessible by the miner.¹⁷⁹ Among them, 16 cited the same reasons for preserving the consent requirement under the research exception (*ie*, that misuse of personal data would lead to irreparable harms) as equally justifying why the exception should be subject to rights of reservation through such means.¹⁸⁰

VI. Comparing cost differentials – An economic paradigm

65 With economic considerations being paramount in the law of copyright in Singapore, we also wanted to explore the domain of law and economics to identify legal reforms that are economically efficient and can encourage socially desirable behaviour among DM stakeholders in the long run.

As mentioned, a hybrid legal defence strategy (*ie*, pleading both fair use defence and specific DM exception) is often unworkable due to the incompatibility between both defences when applied to the DM context. By forcing the hand of litigants to choose between the two, this may prevent the maximisation of aggregate welfare if reliance on the costlier option later becomes predominant. Therefore, we specifically conducted an economic analysis to discern which of the two defences is the costlier approach in the long run to better inform related stakeholders.

67 The concept of a specific exception or an open-ended fair dealing/ fair use defence is not new to Singapore's copyright legislation. From an economic vantage point, the specific exception can be described as a *rule*, while the open-ended defence can be broadly characterised as a *standard*; the former usually associated with detailed close-ended provisions, while the latter with more general and open-ended drafting.¹⁸¹ As a further

¹⁷⁷ Interview with A1 and B1.

¹⁷⁸ Interview with A1 and B1.

¹⁷⁹ Interview with all Category A and B respondents including A1 to A11 and B1 to B11.

¹⁸⁰ Interview with A2, A3, A4, A6, A7, A8, A9, A10, B1, B2, B3, B4, B5, B7, B8 and B10.

¹⁸¹ See Michael Handler & Emily Hudson, "Fair Use as an Advance on Fair Dealing? Depolarising the Debate" in *The Cambridge Handbook of Copyright Limitations* (cont'd on the next page)

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

comparison, rules set out in advance the legal consequences of a particular behaviour of set of facts (*ex post*), while standards provide guidance regarding the appropriate legal response but leaves that determination to a judge or adjudicator (*ex ante*).¹⁸² For example, a "rule" might render it an offence to drive above 60 kilometres per hour on a particular stretch of road, whilst a "standard" might prohibit driving at an "excessive" speed.

68 Currently, the rules and standards literature does not posit any one form of approach as preferable to another.¹⁸³ Where they tend to converge, however, is espousing a similar objective that *the* preferable defence is one that is adaptable to *new technologies* without the need for legislative intervention. The Australian Law Review Committee ("ALRC") notes that this not only reduces ongoing rounds of legal reform, but also prevents the legislature from having to predict in advance the precise uses that would come within the scope of an unremunerated exception.¹⁸⁴ Today, in a world where the uses of DM technologies have been rapidly evolving over time, it pays to compare how this would affect the relative costs incurred under a rules or a standards-based approach, to identify the approach that best encourages socially desirable behaviour, *ie*, an increase in public benefit, over time as being the less costly of the two.

69 Where technology changes rapidly under a rules-based approach, this can incur error costs when rules fail to adapt early on to the new environment known as type 1 error costs,¹⁸⁵ which represents the risks of chilling socially desirable behaviour as a result of decisions made by judges and adjudicators based on imperfect information. More prominently, Louis Kaplow at Harvard Law School argues that such costs are long drawn due to the substantial lag before rules can be amended

© 2021 Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

and Exceptions (Shyamkrishna Balganesh, Ng-Loy Wee Loon & Haochen Sun eds) (Cambridge University Press, 2021) at p 140. See also Isaac Ehrlich & Richard A Posner "An Economic Analysis of Legal Rulemaking" (1974) 3(1) *Journal of Legal Studies* 258.

¹⁸² See Michael Handler & Emily Hudson, "Fair Use as an Advance on Fair Dealing? Depolarizing the Debate" in *The Cambridge Handbook of Copyright Limitations and Exceptions* (Shyamkrishna Balganesh, Ng-Loy Wee Loon & Haochen Sun eds) (Cambridge University Press, 2021) at p 140.

¹⁸³ Emily Hudson, "Implementing Fair Use in Copyright Law: Lessons from Australia" (2013) 25 Intellectual Property Journal 202 at 226. Hudson argues that "[f]air use should not be seen as inevitably superior to specific exceptions, or the endpoint of a mature copyright system ... there will be times when a simple rule is superior; instances when a multi factor standard is preferable; and still other times when the best approach is a well-drafted complex rule".

¹⁸⁴ Australian Law Reform Commission, *Copyright and the Digital Economy* (Report No 122, 2013) at para 6.17.

¹⁸⁵ Deloitte Access Economics, "Copyright in the Digital Age: An Economic Assessment of Fair Use in New Zealand" (February 2018) at p 55 https://www2.deloitte.com/ nz/copyright-digital-age (accessed 11 September 2021).

legislatively.¹⁸⁶ International experience to this effect can be observed in countries such as New Zealand, which did not introduce a specific fair dealing exception for time shifting of television recordings¹⁸⁷ until *more than 20 years* after the recognition of such permissibility by the US Supreme Court under the fair use defence.¹⁸⁸ This reflects the ALRC's warnings that "a confined fair dealing exception will be less flexible and less suited to the digital age than an open-ended fair use exception".¹⁸⁹ Deloitte Access Economics had recently completed a comprehensive study on fair dealing/fair use in Australia and New Zealand, with the observation that "laws which are promulgated as standards rather than rules such as fair use are likely to provide an environment that is more responsive to technological or social change".¹⁹⁰ A conditional probability model that calculates type 1 error costs associated with the rules-based approach is set out in Annex C.

A type 1 error can also be incurred under a standards-based approach, when there is a judicial impetus to retain an "old" range of standards despite a new range being more economically efficient under a new environment altered by technological advances. However, given the breadth of this approach, judges may also mistakenly interpret a new range from the same factors despite no actual change in the environment, incurring type 2 error costs. The conditional probability model depicting both type 1 and type 2 error costs under the standards-based approach is similarly set out in Annex D.

71 Annex E then presents the mathematical modelling of comparing total error costs incurred under both approaches, from which a principal finding emerges: that the *more rapid* the change in technology, the *more economically inefficient* it is to adopt a rules-based approach as it results in greater error costs. In turn, for DM technologies of which its uses continue to rapidly evolve in a multitude of ways, a specific DM exception would likely discourage socially desirable behaviour in the long run. Conversely, a standards-based approach (*ie*, fair use) would be a preferable option for

¹⁸⁶ Louis Kaplow, "Rules versus Standards: An Economic Analysis" (1992) 3 Duke Law Journal 42.

¹⁸⁷ Copyright Act 1994 (No 143) (New Zealand) s 84.

¹⁸⁸ Sony Corp of America v Universal City Studios, Inc 464 US 417 (1984).

¹⁸⁹ Australian Law Reform Commission, Copyright and the Digital Economy (Report No 122, 2013) at para 6.17. See also Deloitte Access Economics, "Copyright in the Digital Age: An Economic Assessment of Fair Use in Australia" (February 2018) at para 5.2 https://www2.deloitte.com/au/en/pages/economics/articles/copyrightdigital-age-google.html> (accessed 11 September 2021).

¹⁹⁰ Deloitte Access Economics, "Copyright in the Digital Age: An Economic Assessment of Fair Use in Australia" (February 2018) at para 5.1 https://www2.deloitte.com/ au/en/pages/economics/articles/copyright-digital-age-google.html> (accessed 11 September 2021).

No part of this document may be reproduced without permission from the copyright holders.

regulating the use of DM technologies as it has a stronger likelihood of accommodating socially desirable DM activities over time. The Deloitte report also concludes that "a standards-based fair use framework can develop over time without incurring the costs, delays and rent-seeking inevitably associated with introducing statutory amendments".¹⁹¹

VII. Summary of findings and recommendations

A. Preference for the open-ended fair use/fair dealing defence

72 The key findings regarding the defences under the Copyright Act point to the open-ended fair dealing/fair use defence as the preferable approach. First, the doctrinal analysis shows that the factors under the open-ended fair dealing defence under s 35(2) of the current Copyright Act can be adequately applied to DM. Second, the findings from the empirical study also indicate a general preference among owners and miners for the open-ended fair use defence. Third, the economic analysis indicates that a specific exception presents higher risks of chilling socially desirable behaviour for a rapidly changing environment like DM (the proven assumption that $ErrorCost_{Rules} > ErrorCost_{Standards}$ with an increase in p), not forgetting that such costs can be prolonged since there will be a substantial lag before legislative amendments can take place to accommodate new advances in technology (Kaplow's observation).¹⁹² In a way, the fact that the proposed Copyright Act 2021 is only being legislated into law two years after Proposal 8 of the Report was introduced is emblematic of this implementation lag.¹⁹³

73 The findings under an economic analysis also have profound implications in the Singapore context, since moving ahead, miners (prospective defendants in the suit) would likely be forced to make a practical choice between pleading the open-ended fair use or specific computational data analysis exception. Hence, should the computational data analysis exception become the more favoured defence of the two to

¹⁹¹ Deloitte Access Economics, "Copyright in the Digital Age: An Economic Assessment of Fair Use in Australia" (February 2018) at para 5.1.2 https://www2.deloitte.com/au/en/pages/economics/articles/copyright-digital-age-google.html> (accessed 11 September 2021).

¹⁹² Louis Kaplow, "Rules *versus* Standards: An Economic Analysis" (1992) 3 *Duke Law Journal* 42.

¹⁹³ The Copyright Bill (Bill 17 of 2021) was only introduced in February 2021 and was still subject to further revisions, and public consultation on the Bill only ended on 8 April 2021, which was slightly over two years after the Copyright Review Report was released back on 17 January 2019. The Bill was finally tabled in Parliament for the first reading on 6 July 2021.

No part of this document may be reproduced without permission from the copyright holders.

plead in the future, this could diminish socially desirable DM activities in Singapore in the long run given the higher error costs associated with this rules-based approach. To make matters worse, such a chilling effect would also be lasting due to the implementation lag. Contrastingly, given the lower error costs under a standards-based approach, the open-ended defence would be better able to respond to changing environments in DM to accommodate socially desirable behaviour, and thus is more *future-proof* in comparison.

B. Clarifying the meaning of "lawful access" under s 244(d) of the proposed Copyright Act 2021

Nonetheless, as it stands from the proposed Copyright Act 2021, Singapore is likely to concurrently pursue a rules-based approach by introducing a specific exception for "computational data analysis" under s 244 (*ie*, the specific DM exception equivalent of Proposal 8). We have a number of cautionary observations in this regard.

It is noteworthy that the proposed s 244 implicitly acknowledges 75 the fact-expression dichotomy and extends protection to DM of copyrighted works. Since s 243(a) defines "computational data analysis" as "using a computer program to identify, extract and analyse information or data from the work" [emphasis added], it is clear that this provision recognises scenarios where the reproduction of *works* (*ie*, expressions) using web robots is merely a means to analyse the data or information contained within these works (ie, facts). This happens especially when data or information is not in alpha-numeric format such that the reproduction of such "facts" cannot be independently collected without the prior reproduction of the work itself. For instance, in certain scenarios when artistic works like paintings are being mined for analysis, a process known as region-based segmentation requires reproducing and downloading the image into the coding software before the algorithm can analyse the data/information within by separating objects in the picture into different categories based on a pre-defined threshold value.¹⁹⁴ Section 244 thus accommodates new and more intricate methods of DM.

76 However, to avoid infringement liability under s 244, miners must further show that they satisfy the conditions under s 244(2). As explained earlier, the "lawful access" requirement that has been transposed from

¹⁹⁴ Pulkit Sharma, "Computer Vision Tutorial: A Step-by-Step Introduction to Image Segmentation Techniques (Part 1)" (1 April 2019) https://www.analyticsvidhya. com/blog/2019/04/introduction-image-segmentation-techniques-python/> (accessed 17 May 2021).

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

Proposal 8 to s 244(2)(*d*) is currently the most contentious requirement,¹⁹⁵ of which the present illustrations listed under s 244 of the proposed Copyright Act 2021 are also of *limited* usefulness as they only clarify that the breach of terms of use or the circumvention of paywalls are unlawful.¹⁹⁶

⁷⁷ In this regard, the issue regarding the lack of clarity under the "lawful access" requirement persists since these illustrations fail to address a number of other important scenarios. As identified earlier, a key scenario is when web robots maliciously bypass REPs. At first glance, this has been directly addressed in the European Union ("EU") by Art 4 of the Directive on Copyright in the Digital Single Market ("DSM Directive"),¹⁹⁷ which exempts reproductions of works and subject matter that have been lawfully accessed for the purposes of DM from legal liability. Furthermore, Art 4(3) clarifies that the reproduction of online content that has been reserved by rights holders through "*machine-readable means* in the case of content made publicly available online" [emphasis added], such as metadata,¹⁹⁸ shall constitute *unlawful*

198 DSM Directive Recital 18. Recital 18 further defines metadata as data providing information about one or more aspects of *other* data.

No part of this document may be reproduced without permission from the copyright holders.

¹⁹⁵ Other non-contentious requirements under s 244 of the Copyright Bill (Bill 17 of 2021) under a DM context would include:

⁽a) Proving that the reproduction of the original was made for the purpose of either computational data analysis itself, or preparing the work for computational data analysis, which is easily satisfied in a DM context (s 244(2)(a) read together with s 244(2)(b)).

⁽b) That the miner does not supply the reproduction to any person other than for the purpose of verifying the results of the computational data analysis or collaborate research and study (s 244(2)(c)). This is likely a straightforward question that can easily be answered in the affirmative or negative when applied to the facts.

⁽c) Establishing that the miner does not know that the original is an infringing copy (s 244(e)(ii)(A)), *or* that the use of the infringing copy is necessary for the purpose for which the miner was carrying out the computational data analysis and for nothing other than this purpose (s 244(e)(iii)). The latter requirement is more easily established, since it is not too different from ss 244(2)(a) and 244(2)(b), with the additional requirement that it was not to be used for other purposes than computational data analytics.

¹⁹⁶ Copyright Bill (Bill 17 of 2021) s 244, Illustrations.

¹⁹⁷ Paragraph 1 of Article 4 of the Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC ("DSM Directive") states: "Member State shall provide for an exception or limitation to the rights ... for reproductions and extractions of *lawfully accessible* works and other subject matter for the purposes of text and data mining" [emphasis added]. For a critique of this provision, see Benjamin Sobel, "A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning" in *Artificial Intelligence and Intellectual Property* (Jyh-An Lee, Reto M Hilty & Kung-Chung Liu eds) (Oxford University Press, 2021) at p 221.

access. Essentially, this forms the only exception in the DSM Directive that can be overridden by contract.¹⁹⁹ Yet, their EU intellectual property policies are known for their strong commitment to the promotion and protection of human rights,²⁰⁰ which are premised on what is often an implicit balance between the rights of creators and the rights/interests of the wider society. Relatedly, all EU member states²⁰¹ are concurrent State Parties to the International Covenant on Economic, Social, and Cultural Rights ("ICESCR"),²⁰² which is the major international human rights instrument addressing human rights issues. Article 15 of the ICESCR states that countries that have ratified or acceded to this instrument "recognise the right of everyone" not only "to enjoy the benefits of scientific progress and its applications",203 but also "to benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author" [emphasis added].²⁰⁴ More prominently, the EU member states have effectively undertaken "to respect the freedom indispensable for scientific research and creative activity".²⁰⁵ Naturally, in implementing and ratifying the DSM Directive, EU member states are obligated to comply with these human rights

See also Jonathan Griffiths & Luke McDonagh, "Fundamental Rights and European IP Law: The Case of Art 17(2) of the EU Charter" in *Constructing European Intellectual Property: Achievements and New Perspectives* (Christophe Geiger ed) (EIPIN and Edward Elgar Publishing, 2013) at p 75 (the authors are also cognisant of the criticisms of a lack of coherence amongst Member States). See, *eg*, Christophe Geiger, "The Constructing European Intellectual Property in the European Union: Searching for Coherence" in *Constructing European Intellectual Property: Achievements and New Perspectives* (Christophe Geiger ed) (EIPIN and Edward Elgar Publishing, 2013) at p 5.

¹⁹⁹ Article 7(1) of the DSM Directive states that "[a]ny contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable".

²⁰⁰ See, for example, Gráinne de Búrca, "The Road not Taken: The European Union as a Global Human Rights Actor" (2011) 105 *American Journal of International Law* 649 at 668:

[[]R]espect for fundamental rights – inspired by the common constitutional traditions of the member states and the international human rights treaties on which they collaborated – was declared a general principle of Community law, and the ECJ would henceforth entertain claims that such rights had been adversely affected by Community acts and policies.

²⁰¹ The EU countries are: Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain and Sweden.

²⁰² International Covenant on Economic, Social and Cultural Rights (16 December 1966), 993 UNTS 3 (entry into force 3 January 1976) ("ICESCR").

²⁰³ ICESCR Art 15.1.

²⁰⁴ ICESCR Art 15.1(c).

²⁰⁵ ICESCR Art 15.3.

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

standards under the ICESCR, as well as under the European Convention on Human Rights.²⁰⁶

78 The human rights norms that have influenced the jurisprudence of the DSM Directive arguably distinguishes the DSM Directive from Singapore's intellectual property paradigm that is underpinned by economic considerations in multiple ways. First, by ensuring everyone a right to the benefits of science and technology, the ICESCR enables easier access for both individuals and communities, which cannot be achieved without robust government policies that actively calibrates the balance of interests away from creators and in favour of the individuals and collective at large, even at the expense of going against market forces. Second, according everyone a right to the protection of their "moral and material interests" would similarly demand greater intervention to safeguard individuals from possible harmful effects of scientific and technological development. All these considerations could potentially render the DSM Directive an inappropriate comparison or benchmark for the development of Singapore's copyright law.

For Art 4(3) of the DSM Directive to be taken as a valid reference 79 point for further improvements to s 244 of the proposed Copyright Act 2021, greater justifications are thus required to show that Art 4 of the DSM Directive and Singapore's copyright regime are more similar rather than distinguishable. Conceptually, Art 4 of the DSM Directive can be described as an "opt-out" consent regime - the owner is assumed to have granted consent for the public to access its web servers that are connected to a public network where mining activities can freely take place, unless owners opt-out from this implied consent arrangement through "content made publicly online" under Art 4(3). Importantly, this interpretation is persuasive as it finds support from academic commentators such as Gove N Allen, who argues that uploading one's works to web servers that are connected to a public network in which the use of automated retrieval mechanisms is pervasive, without placing any restrictions on its reproduction, reasonably indicates that the owner has given "implied consent" for the public to access to the author's works.²⁰⁷ Should such a conception be adopted, it follows that any indication otherwise by the

²⁰⁶ Convention for the Protection of Human Rights and Fundamental Freedoms (4 November 1950), (entry into force 3 September 1953).

²⁰⁷ Gove N Allen, Dan L Burk & Gordon B Davis, "Academic Data Collection in Electronic Environments: Defining Acceptable Use of Internet Resources" (2006) 30(3) *Management Information Systems Quarterly* 601 at 606 (the authors argue that for webpages that fail to provide a terms of service agreement or Robots Exclusion Protocol, one might reasonably infer that an owner grants "implied consent" for the public to access its web servers since they are connected to a public network in which the use of automated retrieval mechanisms is pervasive).

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

author would likely rebut the presumption that the author has given "implied consent".

80 Compared to the Copyright Act, where its scope covers *works* that attract copyright protection (*ie*, excluding facts such as personal data), the harms related to the mining of works are less severe relative to a misuse of personal data, and this would bolster the case for adopting an "opt-out" regime under s 244 of the proposed Copyright Act 2021.²⁰⁸ Such an approach also fits well within the economic goals of Singapore's copyright laws by simplifying the process for miners to engage in the mining of online works that are not subject to reservations made by individuals to the contrary.

81 In this regard, in line with the drafting of s 244 that more prominently relies on the use of illustrations, we recommend the adoption of similar illustrations to Art 4(3) of the DSM Directive, and to incorporate an "opt-out" regime under s 244 of the proposed Copyright Act 2021. Specifically, this clarifies that the legal status of REPs should be equivalent to paywalls or TOS agreements; such that the malicious act of bypassing REPs would constitute *unlawful* access and therefore a failure to satisfy the requirement under s 244(2)(d).

82 In the same vein, where there is *absolutely no attempt* by the owner to circumscribe their rights to their works (*ie*, either directly through TOS agreements or indirectly through paywalls), the owner should be taken to have granted "implied consent" for the public to access and use their works for any purpose. Accordingly, where evidence showing that attempts by the owner to circumscribe rights in their works is *completely non-existent*, an illustration should be inserted to clarify that the mining of such unrestricted works should constitute *lawful* access of the work.

83 Finally, further issues regarding the "lawful access" requirement have also been enumerated by stakeholders in the empirical study and these need to be addressed in the future. For instance, it was raised that this requirement would be difficult to apply to TOS agreements that may or may not be *accessible* to the user.²⁰⁹ This likely involves a fact-specific

²⁰⁸ A distinction can be justified on grounds that a misuse of personal data has the potential to cause irreparable harm to the identifiable individual, thereby favouring an "opt-in" regime under which the consent obligation is more onerous to discharge.

²⁰⁹ Interview with B5. Note that the element of *accessibility* does not affect Robots Exclusion Protocols ("REPs"), paywalls or firewalls since these mechanisms are easily standardised (*eg*, the only way to check whether the owner has reserved his or her rights through REPs is to type in "robots.txt" after the web domain name). The positioning of terms of service agreements on webpages, on the other hand, can be varied, thereby affecting its accessibility to users in different ways.

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

exercise since much would depend on the *type* of TOS agreement used and the *location* where the agreement was displayed on the webpage. TOS agreements such as click-wrap agreements are easily more accessible to the user when it appears in a dialog box or pop-up window the moment the webpage is visited, as compared to browse-wrap agreements that are hidden in obscure locations on the webpage. In turn, the greater the ease of accessibility of the agreement, the more likely that the miner would have been privy to the terms of use of the site, which supports the finding that such works, if mined, would constitute *unlawful* access. Naturally, the converse is also true – the less accessible the agreement, the less likely it can be imputed that the miner is privy to the terms of use, favouring the argument that the mining of the work still constitutes *lawful* access under s 244(2)(d).

84 Illustrations, if any, should only reflect *clear* situations where the use of such agreements are *easily accessible* such that the mining of related works would *most likely* constitute unlawful access. A fitting illustration should thus include click-wrap agreements appearing in a dialog box that further requires the activation of an "I Agree" button to strongly signal the user's acceptance of restrictions imposed by the owner of the webpage. Understandably, the non-binding nature of illustrations will also provide leeway for the litigants to distinguish their particular circumstance from such illustrations; for example, the wording of the click-wrap agreement is unclear or too small for the user to read such that it would be unreasonable to assume that the user can be privy to or have accepted such restrictions.

85 In summary, given the multiple permutations of these examples, specifying all these situations would be impossible, and instead would provide an *illusion of certainty* where certainty cannot be achieved *ex ante*. However, in addition to the two illustrations in the computational data analysis exception in s 244, we would like to propose three more illustrations as follows:

(a) X does not have lawful access to the original if access to the original has been restricted by Y through machine-readable means in the case of content made publicly available online, such as metadata.

(b) X will have lawful access to the original if access to the original has been restricted by Y through any means, including but not limited to, paywalls or any machine-readable means in the case of content made publicly available online.

(c) X does not have lawful access to the original if X accessed the original in breach of the terms of use of a database, especially when specified under a click-wrap agreement that requires the

No part of this document may be reproduced without permission from the copyright holders.

end-user to manifest his or her assent by clicking an "agree" button on a dialog box.

VIII. Conclusions

86 While the authors have argued for having only an open-ended general fair use defence to deal with DM, we are cognisant that the proposed Copyright Act 2021 which contains a specific computational data analysis exception in s 244 is likely to be enacted without significant changes before the end of the year. Our modest hope is for the inclusion of more illustrations in s 244. We have also contended that the research exception under the PDPA should remain unchanged as it adequately balances the competing interest of miners who wish to continue engaging in DM activities, and individuals who wish to protect their personal data.

87 We recognise the limitations in the economic analysis by not discussing the relative *benefits* between (a) the open-ended fair dealing defence *versus* a specific exception, and (b) whether the research exception should apply also to the collection of personal data, and not just to use and disclosure, and we believe that further studies could be conducted in this regard. Another limitation is the small sample size in the survey. Nonetheless, we were careful to avoid making any generalisations about DM stakeholders in Singapore, and instead focused on obtaining a comprehensive and detailed-oriented account from each DM stakeholder to supplement our legal analysis.

We have attempted a calculation of cost differentials between the open-ended defence and the specific exception which revealed that a specific exception might raise potential costs, which may chill socially desirable DM activities in the long run. Conversely, the open-ended fair use defence can better respond to changing environments in DM to accommodate behaviour that promotes the public benefit. It should be noted that the "lawful access" requirement in specific computational data analysis exception is ambiguous, and it ought to be better clarified.

89 In conclusion, this article has demonstrated the need to ensure that our copyright laws accord with the purpose for which all copyright laws are enacted – namely the promotion of creativity and innovation for the public good. As the Deloitte reports have pointed out, the reality is that with the accelerating pace of technological change, legislative adjustment to specific exceptions has proven neither timely nor effective.²¹⁰ Data analytics is an inevitable technological development of the 21st century, and more specifically predictive analytics can find patterns contained within data in order to detect risks and opportunities, and are applicable to a panoply of activities in the finance, healthcare, retailing, pharmaceuticals, automotive, aerospace and manufacturing industries. This inevitably requires a fair balance to be struck between the remuneration for authors, and the access that should be granted to other users to enable them to copy these works in order to create new ones in the advancement of the public good.

²¹⁰ Deloitte Access Economics, "Copyright in the Digital Age: An Economic Assessment of Fair Use in Australia" (February 2018) at para 5.1.2 <https://www2.deloitte.com/ au/en/pages/economics/articles/copyright-digital-age-google.html> (accessed 11 September 2021); Deloitte Access Economics, "Copyright in the Digital Age: An Economic Assessment of Fair Use in New Zealand" (February 2018) at paras 5.1 and 5.2.2 <https://www2.deloitte.com/nz/copyright-digital-age> (accessed 11 September 2021).

^{© 2021} Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

Annex A

Interview respondents

| Category A: Miners | | | |
|--------------------|--|-------------------|--|
| Identifier | Type of mining activities | Date of Interview | |
| A1 | For internships in a FAANG (Facebook, | 18 August 2020 | |
| | Amazon, Apple, Netflix, Google) company | | |
| A2 | For internships in a FAANG (Facebook, | 18 August 2020 | |
| | Amazon, Apple, Netflix, Google) company | | |
| A3 | For internships in a small- | 21 August 2020 | |
| | medium enterprise | | |
| A4 | Teaches mining to students in a university | 22 August 2020 | |
| A5 | Teaches mining to students in a university | 23 August 2020 | |
| A6 | Teaches mining to students in a university | 16 September 2020 | |
| A7 | For work-related purposes in a small- | 14 October 2020 | |
| | medium enterprise | | |
| A8 | For work-related purposes in a | 10 December 2020 | |
| | government sector | | |
| A9 | Teaches mining to students in a university | 13 December 2020 | |
| A10 | Teaches mining to students in a university | 5 January 2021 | |
| A11 | For work-related purposes in a | 12 March 2021 | |
| | government sector | | |

| Category B: Owners | | | |
|--------------------|--|-------------------|--|
| Identifier | Type of data owned | Date of Interview | |
| B1 | Runs a start-up that handles personal data | 5 December 2020 | |
| B2 | Deals with data collected by hospitals for | 7 December 2020 | |
| | work-related purposes | | |
| B3 | Deals with data collected from international | 18 December 2020 | |
| | organisations for work-related purposes | | |
| B4 | Creates content intended for | 28 December 2020 | |
| | commercial exploitation | | |
| B5 | Runs a start-up that handles personal data | 5 January 2021 | |
| B6 | Runs a start-up that handles personal data | 8 January 2021 | |
| B7 | Creates content intended for | 8 January 2021 | |
| | commercial exploitation | | |
| B8 | Runs a start-up that handles copyrightable | 1 February 2021 | |
| | material and personal data | | |
| B9 | Runs a start-up that handles personal data | 8 February 2021 | |
| B10 | Runs a start-up that handles personal data | 10 February 2021 | |
| B11 | Works for a company and handles | 12 March 2021 | |
| | copyrightable material and personal data | | |

 $^{\odot}$ 2021 Contributor(s) and Singapore Academy of Law.

No part of this document may be reproduced without permission from the copyright holders.

Annex B

Interview questions

| Category A: Miners | | | |
|--------------------|---|---|--|
| Classification | | Questions | |
| Introductory | 1 | What kind of data mining ("DM") activities are you | |
| Questions | | involved in? | |
| | 2 | Do you believe that there is a need to regulate DM? | |
| | 3 | Can you think of other useful DM activities that should | |
| | | be made legitimate? | |
| About the | 4 | Do you think that the current factors under s 35(2) of | |
| Copyright Act | | the Copyright Act can be adequately applied to a given | |
| | | context like DM? | |
| | 5 | (a) Do you think that the current requirements under | |
| | | Proposal 8's specific DM exception is adequate? (b) Do | |
| | | you think this definition of "lawful access" is clear? | |
| | 6 | Between s 35(2) of the Copyright Act and the new DM | |
| | | exception under the Proposal 8, which approach do you | |
| | | find to be preferable? | |
| About the | 7 | Do you think that the research exception should | |
| Personal Data | | continue to retain the consent requirement for the | |
| Protection Act | | collection of personal data? Why? | |
| | 8 | Should miners be allowed to rely on the research | |
| | | exception even when owners have reserved rights to | |
| | | their data to the contrary? | |

| Category B: Owners | | | |
|----------------------------|---|--|--|
| Classification | Questions | | |
| Introductory Questions | What kind of data do you own that could be potentially mined? Do you believe that there is a need to regulate DM? Can you think of other useful DM activities that should | | |
| | be made legitimate? | | |
| About the Copyright Act | 4 Do you think that the current factors under s 35(2) of the Copyright Act can be adequately applied to a given context like DM? | | |
| | 5 (a) Do you think that the current requirements under Proposal 8's specific DM exception is adequate? (b) Do you think this definition of "lawful access" is clear? 6 Between s 35(2) of the Copyright Act and the new DM exception under the Proposal 8, which approach do you | | |
| | find to be preferable? | | |

No part of this document may be reproduced without permission from the copyright holders.

| (2021) 33 SAcLJ | | Copying Right in Copyright Law: Fair Use, Computational Data Analysis and the PDPA | 1079 |
|-----------------|---|---|------|
| | - | | |
| About the | 7 | Do you think that the research exception should | |
| Personal Data | | continue to retain the consent requirement for the | |
| Protection Act | | collection of personal data? Why? | |
| | 8 | Should miners be allowed to rely on the research exception even when owners have reserved rights to their data to the contrary? | |

Annex C

Conditional probability modelling of a rules-based approach

Diagram 1 below illustrates the conditional probability of outcomes under a rules-based approach through a probability tree, together with the expected error costs.



Diagram 1: Probability tree of possible outcomes for rules-based approach

The expected error cost (" $ErrorCost_{Rules}$ ") under a rules-based approach can be expressed as follows:

 $ErrorCost_{Rules} = prob (type \ 1 \ error)$ $= p \times C1$ $= p \cdot C1$

Annex D

Conditional probability modelling of a standards-based approach

Diagram 2 below illustrates the probability tree depicting the conditional probability of outcomes under a standards-based approach through a probability tree.



Diagram 2: Probability tree of possible outcomes for a standards-based approach

The expected error cost ("*ErrorCost*_{Standards}") under a rules-based approach can be expressed as follows:

$$ErrorCost_{Standards} = prob (type 1 error) + prob (type 2 error)$$

= $[p \times (1 - r) \times C1] + [(1 - p) \times w \times C2]$
= $(1 - r) \cdot p \cdot C1 + (1 - p) \cdot w \cdot C2$

Annex E

Mathematical modelling comparing the total error costs associated with rules-based and standards-based approaches

From the findings in Annexes C and D, the following mathematical modelling assumes that the error costs incurred under a rules-based approach is greater than that under a standards-based approach (*ErrorCost*_{Rules} > *ErrorCost*_{Standards}). The resulting inequality would emerge:

No part of this document may be reproduced without permission from the copyright holders.

$$\begin{split} & ErrorCost_{Rules} > ErrorCost_{Standards} \\ & \rightarrow p \cdot C1 > (1 - r) \cdot p \cdot C1 + (1 - p) \cdot w \cdot C2 \\ & \rightarrow (1) \cdot p \cdot C1 - (1 - r) \cdot p \cdot C1 > (1 - p) \cdot w \cdot C2 \\ & \rightarrow [1 - (1 - r)] \cdot p \cdot C1 > (1 - p) \cdot w \cdot C2 \\ & \rightarrow r \cdot p \cdot C1 > (1 - p) \cdot w \cdot C2 \\ & Inequality \rightarrow \frac{r}{w} > \frac{(1 - p)}{p} \cdot \frac{C2}{C1} \end{split}$$

Consequently, *for this inequality to hold true* (and thus proving the assumption that underlies it), Diagram 3 below breaks down three ratios from this inequality that are relevant to this analysis.

| Assumption $\rightarrow ErrorCost_{Rules} > ErrorCost_{Standards}$ | | | | | |
|--|---|----------|---|--|--|
| Ratio | Analysis | I: tl | s it possible to estimate he value of these ratios? | | |
| $\frac{r}{w}$ | <u><i>r</i></u> <i>w</i> , which reflects the reliability of a court's ability to correctly identify environmental change under the standards-based approach, must be naturally <i>higher</i> . | × | No, due to a lack of local case law to analyse the ability of local judges to identify environmental changes correctly. | | |
| $\frac{(1-p)}{p}$ | (1-p) <i>p</i> , which reflects the stability of the environment, must preferably be <i>lower</i> . By extension, <i>p</i> , which reflects the likelihood of the environment changing, should be <i>greater</i> . | | Yes, because transformation (<i>p</i>) is rapid in DM, hence it is reasonably predictable that the value of <i>p</i> would be greater. | | |
| <u>C2</u> <u>C1</u> | $\frac{C2}{C1}$, which represents the relative costs between a type 2 as against a type 1 error, must ideally be <i>lower</i> . Preferably, this entails that the absolute cost of C1 (<i>ie</i> , type 1 error) should be higher than C2 (<i>ie</i> , type 2 error). | × | No, since it is difficult to quantify C1 and C2, which involves a calculation of various costs that cannot be undertaken within a year- long study. | | |

Diagram 3: Breakdown of variables within the inequality derived from the comparison of costs between a rules-based and a standards-based approach.

Diagram 3 shows that only the value of *p* is discernible in the DM context. This in turn influences the value of $\frac{(1-p)}{p}$, which weighs the probability of the environment staying the same (*p*) as against the probability of

a shift towards a new environment (1 - p). Given that p and (1 - p) must necessarily add up to 1, an increase in p would concurrently reduce the value of (1 - p), thus decreasing the value of $\frac{(1-p)}{p}$. Since the value of $\frac{(1-p)}{p}$ must preferably be lower for the inequality to hold true, this would therefore prove the underlying assumption behind this inequality: that the error costs incurred under a rules-based approach is greater than that under a standards-based approach with (*ErrorCost*_{Rules} > *ErrorCost*_{Standards}) in an environment where digital transformation is rapid (increase in p).