

8. Non-consumptive Use

Contents

Summary	155
Defining ‘non-consumptive’ use	155
Caching and indexing and network-related functions	157
Current law	158
Criticisms of the current exceptions	158
International comparisons	160
The need for greater certainty	161
Infringement concerns and current practices	162
Interaction with safe harbour review	162
Text and data mining	164
Current law	165
International comparisons	165
Licensing solutions	166
Facilitating research and study	168
Non-consumptive uses and fair use	169
Caching and indexing	169
Data and text mining	170

Summary

8.1 This chapter considers ‘non-consumptive’ uses of copyright material. This captures uses which do not trade on the underlying creative and expressive purpose of the material. Examples of non-consumptive uses include caching and indexing by search engines, and possibly text and data mining.

8.2 The ALRC proposes that the fair use exception proposed in Chapter 4 should be used to determine whether activities such as caching and indexing, or data and text, mining constitute infringement. If fair use is not enacted, the *Copyright Act 1968* (Cth) should be amended to provide for a new fair dealing exception for non-consumptive use.

8.3 The current exceptions in the *Copyright Act* that relate to temporary reproductions should be repealed.

Defining ‘non-consumptive’ use

8.4 In 2011, the Hargreaves Review recommended that the UK implement an exception for ‘non-consumptive use’, which was defined as use of a work enabled by

technology which does not trade on the underlying creative and expressive purpose of the work:

The idea is to encompass the uses of copyright works where copying is really only carried out as part of the way technology works. For instance, in data mining or search engine indexing, copies need to be created for the computer to analyse; the technology provides a substitute for reading all the documents ... that these new uses happen to fall within the scope of copyright regulation is essentially a side effect of how copyright has been defined, rather than being directly relevant to what copyright is supposed to protect.¹

8.5 The digital age has seen the emergence of many ‘copy-reliant technologies’ such as search engines, which copy expressive works for non-expressive aims.² US Professor Matthew Sag has written that:

because expressive communication to the public implicitly defines and limits the extent of the copyright owner’s exclusive rights, acts of copying that do not communicate the author’s original expression to the public do not generally constitute copyright infringement.³

8.6 The idea that certain non-consumptive or non-expressive uses ought not to be protected by copyright can be traced to a fundamental distinction in copyright law: that between ideas and expression.⁴ Australian and overseas courts have reiterated that copyright law does not exist to protect facts or information, but the expression of ideas and information.⁵ Sag notes that the ideas/expression distinction is central to balancing the interest of authors in preventing exploitation of their works and society’s interest in the free flow of ideas, information and commerce:

Subsequent authors may not compete with the copyright owner by offering her original expression to the public as a substitute for the copyright owner’s work, but they are free to compete with their own expression of the same facts, concepts and ideas.⁶

8.7 This distinction is also drawn by Professor Kathy Bowrey, although she refers to the term ‘culturally meaningful uses’:

The economic logic of copyright is tied to a cultural logic. That a use such as caching can be assigned an economic value does not justify a copyright return. A copyright owner’s right is *a right to control culturally meaningful uses*, not every use of a work. To maintain the integrity and consistency of copyright’s logic, dealings with mere data should be distinguished from culturally meaningful dealings with copyright works and subject-matter.⁷

1 I Hargreaves, *Digital Opportunity: A Review of Intellectual Property and Growth* (2011), 47.

2 M Sag, ‘Copyright and Copy-Reliant Technology’ (2009) 103(4) *North Western University Law Review* 1607, 1608.

3 *Ibid*, 1609.

4 A Stewart, P Griffith and J Bannister, *Intellectual Property in Australia* (4th ed, 2010), 154.

5 See eg, *Ice Tv Pty Ltd v Nine Network Australia Pty Ltd* (2009) 239 CLR 458; *Feist Publications, Inc., v Rural Telephone Service Co.*, (499 U.S. 340, 1991).

6 M Sag, ‘Copyright and Copy-Reliant Technology’ (2009) 103(4) *North Western University Law Review* 1607, 1629.

7 K Bowrey, *Submission 94*.

8.8 Two types of uses identified in the Issues Paper are discussed below. These relate to caching, indexing and other internet functions; and data and text mining. In both instances, the ALRC considers that such uses should be considered under the fair use exception. Alternatively, if fair use is not enacted, they should be considered under a new fair dealing exception for ‘non-consumptive’ use.

Caching and indexing and network-related functions

8.9 Internet service providers, search engines, web hosts and other internet intermediaries rely on indexing and caching for their efficient operation. For example, Google’s search engine works by using automated web crawlers that find and make copies of websites on the internet. These copies are then indexed and stored on its cache.⁸ When a user enters a search query, Google uses the cached version to judge if the page is a good match for the query, and displays a link to the cached site.⁹

8.10 Caching improves the internet’s performance by allowing search engines to quickly retrieve cached copies on its server, rather than having to repeatedly retrieve copies from other servers. It is also helpful when the original page is not available due to internet traffic congestion, an overloaded site, or if the owner has recently removed the page from the web.¹⁰

8.11 One Australian internet service provider submitted that approximately 70% of the traffic it delivers to customers is from overseas, and submitted that:

in the event that we are able to cache files locally, we can improve the delivery process, ensure error-free delivery and reduce the time required to download and view content. In our experience, by using caching, transmission overheads can reduce to 1% of what they otherwise would be, without caching.¹¹

8.12 Google and Yahoo!7 made similar comments in their submissions on the value of ‘system-level caching’ to the efficient delivery of internet services to Australians.¹²

8.13 The Australian Broadcasting Corporation (ABC) noted that caching and indexing are ‘an essential part of the technical delivery process’, without which it would be unable to provide reliable streamed television programming over the internet of a quality acceptable to customers.¹³ Similarly, Screenrights noted in relation to its

8 Caching can be described as the copying and storing of data from a webpage on a server’s hard disk so that the page can be quickly retrieved by the same or a different user the next time that page is requested. Caching can operate at the browser level (eg, stored on a computer’s hard drive and accessed by the browser) or at a system/proxy level by internet intermediaries and other large organisation: see, Webopedia, *Proxy Cache* <www.webopedia.com/TERM/P/proxy_cache.html> at 31 July 2012.

9 Ibid.

10 Google Guide, *Cached Pages* <www.googleguide.com/cached_pages.html> at 30 July 2012. A website can specifically prevent a crawler from accessing parts of their website which would otherwise be publically viewable, by inserting a piece of code called ‘robot.txt protocol’.

11 iiNet Limited, *Submission 186*.

12 Google pointed to a report from Lateral Economics which suggests that total internet traffic demand per month in Australia is around 800 Gbps with around 80% of that traffic coming from the United States, 60% of which is capable of being cached. If this traffic could be cached, it was suggested cost savings of \$45 million a year could be made: Google, *Submission 217*. See also, Yahoo!7, *Submission 276*.

13 Australian Broadcasting Corporation, *Submission 210*.

Enhance Direct TV service that ‘the operation of s 200AAA facilitates proxy caching so as to make the service more technically viable for participating institutions’.¹⁴

8.14 The Australian Digital Alliance and Australian Libraries Copyright Committee (ADA and ALCC) submitted that libraries also rely on caching to ‘improve the speed and scalability of user searches’.¹⁵

Current law

8.15 There is no specific exception in the *Copyright Act* that permits the copying or reproduction of copyright material for the purposes of caching or indexing. However, there are a number of provisions that deal with ‘temporary reproductions’ and one specific section that deals with ‘proxy caching’ by educational institutions.

- Sections 43A and 111A allow for the temporary reproduction of a work, an adaptation of a work or an audio-visual item as part of the ‘technical process of making or receiving a communication’.¹⁶
- Sections 43B and 111B provide that copyright in a work or subject-matter is not infringed by temporary copying or reproduction ‘incidentally made as a necessary part of a technical process’ of using a copy of the work or subject-matter.¹⁷
- Section 116AB allows for the reproduction of copyright material on a system or network controlled or operated by, or for, a ‘carriage service provider’ in response to an action by a user in order to facilitate efficient access to that material by that user or other users.¹⁸
- Section 200AAA allows automated caching by computers operated by or on behalf of an educational institution.

Criticisms of the current exceptions

8.16 A review in 2000 of intellectual property legislation expressed concerns about whether ss 43A and 111A were sufficiently wide to cover proxy caching.¹⁹ The review recommended that if there is evidence that caching is not permitted as an exception,

14 Screenrights, *Submission 215*.

15 ADA and ALCC, *Submission 213*. The State Library of NSW advised that it ‘generates significant quantities of digital content of its analogue collections, and collects current digital content, it will want to be able to temporarily index this material for clients to facilitate searching’: State Library of New South Wales, *Submission 168*.

16 *Copyright Act 1968* (Cth) s 43A deals with a work, or adaptation of a work and s 111A deals with audiovisual items. Neither provision applies if the making of the communication is an infringement of copyright: ss 43A(2), 111A(2).

17 It is suggested that ss 43B and 111B could apply to caching: K Weatherall, *Internet Intermediaries and Copyright: An Australian Agenda for Reform* (2011), Policy Paper prepared for the Australian Digital Alliance, 16.

18 ‘Carriage service provider’ is defined in s 78 of the *Telecommunications Act 1997* (Cth) to include a party who uses infrastructure provided by a licensed carrier to supply carriage services to the public. Only public internet access providers such as Telstra Bigpond are deemed carriage service providers.

19 Intellectual Property and Competition Review Committee, *Review of Intellectual Property Legislation under the Competition Principles Agreement* (2000), 108–113.

then the *Copyright Act* should be amended. For example, the review stated that s 43A could be modified to include:

other works temporarily made merely as an element in and so as to enhance the efficiency of the technical process of making or receiving a communication.²⁰

8.17 Stakeholders suggested a number of problems with the current provisions. The Australian and Competition and Consumer Commission (ACCC) submitted that ss 43 and 111A do not capture the full scope of copying and communication that may be undertaken in performance of caching and indexing functions since:

- it is not clear what constitutes ‘temporary’ reproduction;
- infringement issues may arise where copying might not be considered part of the technical ‘process of making or receiving a communication’; and
- the exception does not exempt communication of copyright material, such as when communicated to a user.²¹

8.18 eBay argued that caching exceptions should recognise the making of persistent copies for the purposes of facilitating communication, analysis and research.²² eBay also submitted that lack of protection for such activities makes Australia an unattractive place to locate facilities that deliver, in particular, data analysis and search services over the internet.²³

8.19 Telstra expressed concerns that the current exceptions do not recognise that, in the digital environment, ‘multiple reproductions and communications may occur’.²⁴

8.20 Optus stressed the importance of caching and indexing to cloud computing. For example, in relation to ‘data centres’ it pointed to estimates that

over 85% of total global data centre traffic within its data centre, and between data centre traffic—that is, data that does not go to the end user. This traffic will involve automatic copying, including for backup and parallel processing purposes. Under the current Australian copyright regime, this may involve infringement of copyright laws.²⁵

8.21 The Law Council of Australia argued that the legal position in relation to caching is ‘confused’ and that ‘it is undesirable to have several overlapping, but distinct provisions aimed at the same basic phenomenon and offering only partial and uncertain protection’.²⁶

20 Ibid, 113.

21 ACCC, *Submission 165*. Similar concerns were expressed by Telstra Corporation Limited, *Submission 222* and Law Council of Australia, *Submission 263*.

22 eBay, *Submission 93*.

23 Ibid.

24 Telstra Corporation Limited, *Submission 222*.

25 Optus, *Submission 183*.

26 Law Council of Australia, *Submission 263*.

International comparisons

8.22 A number of other jurisdictions have specific exceptions that deal with caching and indexing. Article 13 of the European Council's *E-Commerce Directive* provides an exception for caching.²⁷ The UK has a specific exception—mirroring the *E-Commerce Directive*—that allows a provider to cache copyright material so long as the service provider:

- does not modify the information;
- complies with any conditions on access to, and updating of, the information;
- does not interfere with the lawful use of technology to obtain the data or use the information; and
- acts expeditiously to remove or disable access to the material upon obtaining knowledge that the work has been removed at the initial source, access has been disabled, or a court or administrative body has ordered such removal or disablement.²⁸

8.23 A similar exception for caching exists in New Zealand under s 92E of the *Copyright Act 1994* (NZ).

8.24 In Canada, s 31.1 of the *Copyright Act* (Can) provides an exception that allows persons who provide services 'related to the operation of the Internet or another digital network' to telecommunicate or reproduce the work or other subject-matter through the internet or that network.²⁹ The section specifically provides that caching to make a telecommunication more efficient does not, by virtue of that act alone, infringe copyright, and appears to cover both the reproduction and communication of cached material.³⁰

8.25 The Canadian *Copyright Act* also provides that copyright owners are limited to injunctive relief against a provider of an 'information location tool'³¹ found to have infringed copyright by making a reproduction of copyright material, or by communicating that reproduction to the public by telecommunication.³²

8.26 In the United States, caching, indexing and communication of search results may be non-infringing under fair use. For example, in *Field v Google Inc* it was held

27 *Directive 2000/31/EC of the European Parliament and of the Council on certain legal aspects of information society services, in particular electronic commerce, in the internal market* (entered into force on 8 June 2000) ('Directive on electronic commerce').

28 *Electronic Commerce Regulations 2002* (UK) reg 18.

29 *Copyright Act 1985* (Can) s 31 provides that infringement does not occur 'solely by providing those means' for the telecommunication or reproduction.

30 *Ibid* s 31.2 The exception is subject to a number of factors including that the person: does not modify work; ensures that caching is done in accordance with industry practice; and does not interfere with the use of technology that is lawful and consistent with industry practice in order to obtain data on the use or the work or other subject matter: s 31.3.

31 This is defined to mean 'any tool that makes it possible to locate information that is available through the Internet or another digital network': *Copyright Act 1985* (Can) s 41.27(5).

32 Providers must adhere to certain conditions to benefit from this protection.

that Google did not infringe copyright by caching a story that the plaintiff had posted to his website.³³ The court considered that the practice was fair use because, among other things, it was ‘transformative’ in nature and there was no evidence that Google intended to profit from the caching.³⁴ It also considered that Google was able to rely on the US safe harbour provisions for intermediate and temporary storage.³⁵ Similar findings were made in *Parker v Google Inc.*³⁶

The need for greater certainty

8.27 Stakeholders who called for reform suggested that the law concerning caching and indexing should be clarified to achieve greater certainty. A number suggested that caching and indexing should be covered by a fair use exception, noting especially the difficulty in drafting a technology-neutral exception.³⁷ The ADA and ALCC argued that a specific exception ‘will always be playing catch up with intermediary activities’.³⁸ Telstra suggested that network functions should be protected by a fair use exception which recognises that ‘multiple reproductions and communications are likely to occur’. It argued that the approaches taken in other jurisdictions

remain device/technology centric and therefore risk becoming obsolete as digital technology functions continue to evolve ... redrafting based on today’s technical knowledge and standards is likely to render the exemption obsolete in the context of future innovations.³⁹

8.28 The Law Council of Australia submitted that a generally worded, abstract provision aimed at essential technical and internet functions could be adopted, with fairness factors ‘to protect the interest of copyright owners’.⁴⁰

8.29 Optus argued that the current exceptions may not adequately account for the fact that cloud computing will increasingly require copying of data that is not accessed by the end user.⁴¹ Universities Australia argued that as universities and other bodies migrate to cloud based systems the education-specific exception under s 200AAA ‘may well come under challenge’.⁴²

8.30 Others supported a tightly-worded exception. For example, the Arts Law Centre of Australia supported an exception ‘limited to what is necessary for the technical

33 *Field v Google* 412 F Supp 2d, 1106.

34 *Ibid*, 1117–23.

35 *Ibid*, 1123–25.

36 422 F Supp 2d 492, 497.

37 Grey Literature Strategies Research Project, *Submission 250*; Law Council of Australia, *Submission 263*; EFA, *Submission 258*; Telstra Corporation Limited, *Submission 222*; National Library of Australia, *Submission 218*; Google, *Submission 217*; ADA and ALCC, *Submission 213*; Law Institute of Victoria (LIV), *Submission 198*.

38 ADA and ALCC, *Submission 213*.

39 Telstra Corporation Limited, *Submission 222*.

40 Law Council of Australia, *Submission 263*.

41 Optus, *Submission 183*.

42 Universities Australia, *Submission 246*.

functions of search engines and other internet intermediaries and has no detrimental effects on artists'.⁴³

Infringement concerns and current practices

8.31 Other stakeholders argued that there was no practical impediment to caching and indexing in Australia and that the law did not need to be changed.⁴⁴ For example, Australian Film and TV bodies argued that:

Caching and indexing are not new internet functions; in fact, they happen every day. Despite this, no provider of caching, indexing or hyper-linking services, other than the ISP in the unique circumstances of *Cooper v Universal Music* has been found liable by an Australian court for copyright infringement by authorisation.⁴⁵

8.32 Concerns were also raised that providing access to cached versions of websites may prevent access to underlying works and deprive owners of advertising revenue from their website.⁴⁶ For example, the Arts Law Centre of Australia argued that to the extent that internet-related functions 'reproduce copyright works or assist copyright infringers to access works', they can prejudice the ability of artists to earn an income.⁴⁷ The collecting society APRA/AMCOS did not support any exception that permitted the caching of downloads 'tethered to subscription services, or material located behind a paywall'.⁴⁸

8.33 The ABC believed that some of these concerns may be addressed as the industry continues 'to drive the development of more sophisticated means to protect digital content made available online, such as encryption, digital rights management (DRM), progressive downloading and client verification/authenticating systems'.⁴⁹ The availability of technological protection measures (TPMs) counterbalances the risk of any unauthorised access to digital content. Google also submitted that the 'robot.txt' protocol is 'widely accepted and allows website owners to have complete control over whether and how their content is indexed'.⁵⁰

Interaction with safe harbour review

8.34 The safe harbour provisions are found in div 2AA of the *Copyright Act*. The scheme limits the remedies available against 'carriage service providers'—as intermediaries—against copyright infringement that takes place on their systems,

43 The Arts Law Centre of Australia argued that there should be implementation of protocols for take down notices procedures that require ISPs, and internet connection hosts to remove works from caching and indexing services: Arts Law Centre of Australia, *Submission 171*.

44 BSA, *Submission 248*; APRA/AMCOS, *Submission 247*; Foxtel, *Submission 245*; ARIA, *Submission 241*; PPCA, *Submission 240* News Limited, *Submission 224*; Screenrights, *Submission 215*; Australian Film/TV Bodies, *Submission 205*; iGEA, *Submission 192*; TVB (Australia) Pty Ltd, *Submission 124*.

45 Australian Film/TV Bodies, *Submission 205*.

46 NSW Young Lawyers, *Submission 195*.

47 Arts Law Centre of Australia, *Submission 171*.

48 APRA/AMCOS, *Submission 247*. See also CCH Australia Ltd, *Submission 105*. CCH argued that web searches should not allow back door access to pay for view content.

49 Australian Broadcasting Corporation, *Submission 210*.

50 Google, *Submission 217*.

which they do not control, initiate or direct. The scheme currently covers certain activities of carriage service providers including:

- Category A—acting as a conduit for internet activities by providing facilities for transmitting, routing or providing connections for copyright material;⁵¹
- Category B—caching through an automatic process;⁵²
- Category C—storing copyright material on their systems or network;⁵³ and
- Category D—referring users to an online location.⁵⁴

8.35 The Australian Government Attorney-General's Department has issued a Consultation Paper which proposes to extend the application of the safe harbour regime, by replacing the term 'carriage service provider' with 'service provider'.⁵⁵ The new term would cover 'internet service providers' and operators of online services.

8.36 A number of stakeholders suggested that caching and indexing by internet service providers could be dealt with by extending the definition of 'carriage service provider' under the safe harbour provisions rather than through copyright exceptions.⁵⁶

8.37 ARIA submitted that the balance established under the safe harbour scheme should not be undermined by copyright exceptions, and that if some entities are not protected in respect of their caching activities, 'this should be reviewed through the scope of the safe harbour scheme'.⁵⁷ APRA/AMCOS suggested that:

It would be appropriate to consider the exception through the prism of the safe harbour provisions, ensuring that any entity that was able to take advantage of the exception was also constrained by an appropriate mandatory code of practice for the use of the cached material.⁵⁸

8.38 Other stakeholders argued that the safe harbour provisions do not provide a complete answer.⁵⁹ For example, eBay submitted that if the safe harbours were:

extended beyond carriage service providers, it could operate to exempt from monetary damages the infringement of copyright that occurs in the course of providing many services online. However, this half way approach would not overcome the logical difficulty of providing a 'safe harbour' to support copyright infringement. It would be far preferable to substitute a wider but targeted exemption.⁶⁰

51 *Copyright Act 1968* (Cth) s 116AC.

52 *Ibid* s 116AD.

53 *Ibid* s 116AE.

54 *Ibid* s 116AF.

55 Australian Government Attorney-General's Department, *Revising the Scope of the Copyright 'Safe Harbour Scheme'*, Consultation Paper (2011), 5.

56 Music Council of Australia, *Submission 269*; Internet Industry Association, *Submission 253*; APRA/AMCOS, *Submission 247*; Australian Copyright Council, *Submission 219*.

57 ARIA, *Submission 241*, noting that 'any such review should also develop a code of conduct to apply to all ISPs requiring them to take steps to address all types of copyright infringing activities on their networks or through the use of their services'.

58 APRA/AMCOS, *Submission 247*.

59 R Xavier, *Submission 146*; eBay, *Submission 93*.

60 eBay, *Submission 93*.

8.39 Similarly, the Law Council of Australia noted that one approach may be to reform exceptions for natural persons and leave service providers ‘to the protection of (extended) safe harbours’. However, it argued that

this approach has its weaknesses. It is not clear that all entities engaged in caching in particular would be online service providers. Further, the Safe Harbours were not designed to operate as a general set of regulations requiring compliance from all actors engaged in network related activities ... Creating additional exceptions for direct infringement will still leave work for the safe harbours, which protect from both direct and authorisation liability.⁶¹

8.40 Under the Terms of Reference, the ALRC is not to duplicate work being done in relation to a safe harbour scheme. However, in the ALRC’s view, safe harbours need not be used protect ‘internet service providers’ from liability for caching and indexing activities that are not infringement because of fair use.

Text and data mining

8.41 Data and text mining has been defined as ‘automated analytical techniques’ that work by ‘copying existing electronic information, for instance articles in scientific journals and other works, and analysing the data they contain for patterns, trends and other useful information’.⁶² Data and text mining has also been described as ‘a computational process whereby text or datasets are crawled by software that recognises entities, relationships and actions’.⁶³

8.42 The growth of digital technology has seen increasing amounts of data stored in databases and repositories. Use of data and text mining to extract patterns across large data sets and journal articles is becoming more widely used in a number of research sectors, including medicine, business, marketing, academic publishing and genomics.⁶⁴ This type of research has been referred to as ‘non-consumptive’ research, because it does not involve reading or viewing of the works.⁶⁵

8.43 The Terms of Reference refer to the general interests of Australians to ‘access, use and interact with content in the advancement of education, research and culture’. Researchers and research institutions have highlighted the value of data mining in paving the way for novel discoveries, increased research output and early identification of problems.⁶⁶

61 Law Council of Australia, *Submission 263*.

62 UK Government Intellectual Property Office, *Consultation on Copyright* (2011), 80. See also, D Sašo, ‘Data Mining in a Nutshell’ in S Džeroski and N Lavrač (eds), *Relational Data Mining* (2001). Data mining programs are often called data-analytics software.

63 IASTMP, *Submission 200*.

64 R Van Nooren, ‘Text Mining Spats Heats Up’ (2013) 495 *Nature* 295 provides examples of text mining including: linking genes to research, mapping the brain and drug discovery.

65 C Haven, *Non-consumptive research? Text-mining? Welcome to the Hotspot of Humanities Research at Stanford* (2012) <<http://news.stanford.edu/news/2010/december/jockers-digitize-texts-120110.html>> at 22 April 2013; Association of Research Libraries, *Code of Best Practices in Fair Use for Academic and Research Libraries* (2012).

66 UK Government, *Consultation on Copyright: Summary of Responses* (2012), 17.

8.44 The Cyberspace Law and Policy Centre submitted that data mining

has the potential to grant ‘immense inferential power’ to allow businesses, researchers and institutions to ‘make proactive knowledge-driven decisions’. There are significant potential commercial benefits—data mining has the potential to improve business profits by allowing businesses to better understand and predict the interests of customers so as to focus their efforts and resources on more profitable areas.⁶⁷

8.45 At the commercial level, the ability to extract value from data is an increasingly important feature of the digital economy. For example, the McKinsey Global Institute suggests that data has the potential to generate significant financial value across commercial and other sectors, and become a key basis of competition, underpinning new waves of productivity growth and innovation.⁶⁸

Current law

8.46 There is no specific exception in the *Copyright Act* for text or data mining. Where the text or data mining process involves the copying, digitisation, or reformatting of copyright material without permission, it may give rise to copyright infringement.

8.47 One issue is whether text mining, if done for the purposes of research or study, would be covered by the fair dealing exceptions. The reach of the fair dealing exceptions may not extend to text mining if the whole dataset needs to be copied and converted into a suitable format. Such copying would be more than a ‘reasonable portion’ of the work concerned.⁶⁹ Nor is it clear whether copying for text mining would fall under the exception relating to temporary reproduction of works as part of a technical process, under s 43B of the *Copyright Act*, but it seems unlikely.

International comparisons

8.48 The need for a specific text mining exception has been hotly contested in the UK. The Hargreaves Review recommended that the UK Government ‘press at EU level for the introduction of an exception allowing uses of a work enabled by technology which do not directly trade on the underlying creative and expressive purpose of the work’.⁷⁰ One example given of such a use was data mining. The report also recommended that the Government ensure that such an exception cannot be overridden by contract.⁷¹

8.49 In response to the Hargreaves Review, the Business, Innovation and Skills Committee of the UK Parliament did not endorse a specific exception to deal with data mining for research. Rather, it urged the Government to encourage the early

⁶⁷ Cyberspace Law and Policy Centre, *Submission 201*.

⁶⁸ McKinsey Global Institute, *Big Data: The Next Frontier for Innovation, Competition and Productivity* (2011), Executive Summary. It is suggested that big data equates to financial value of \$300 billion (US Health Care); 250 billion Euros (EU Public sector administration); global personal location data (\$100 billion in revenue for service providers and \$700 billion for end users).

⁶⁹ *Copyright Act 1968* (Cth) s 40(5) setting out what is a ‘reasonable portion’ with respect to different works.

⁷⁰ I Hargreaves, *Digital Opportunity: A Review of Intellectual Property and Growth* (2011), 47.

⁷¹ *Ibid.*, 51.

development of models in which ‘licences are readily available at realistic rates to all bona fide licensees’.⁷²

8.50 However, the UK Government has proposed to amend the *Copyright, Designs and Patents Act 1988* (UK) so that ‘it is not an infringement of copyright for a person who already has a right to access the work (whether under a licence or otherwise) to copy the work as part of a technical process of analysis and synthesis of the content of the work for the sole purpose of non-commercial research’.⁷³ The rationale for this exception was that

the copying involved in text and data analytics is a necessary part of a technical process, and is unlikely to substitute for the work in question (such as a journal article). It is therefore unlikely that permitting mining for research will itself negatively affect the market for or value of copyright works. Indeed, it may be that removing restrictions from analytic technologies would increase the value of articles to researchers.⁷⁴

8.51 It was also proposed that a licence could not prevent the use of works under the exception, but may impose conditions of access to a licensor’s computer system or to third party systems on which the work is accessed. Where a TPM prevents a researcher from benefiting from this exception, appeal can be made to the Secretary of State.

8.52 Text and data mining has also been considered in the US in the context of ‘transformative use’. In *The Authors Guild v HathiTrust*, the trial judge found that non-expressive uses such as text searching and computational analysis are fair use and therefore do not infringe the copyright in the underlying material.⁷⁵

Licensing solutions

8.53 A number of stakeholders submitted that there was no impediment to data or text mining in the *Copyright Act*.⁷⁶ Some suggested that data and text mining activities may already be covered under the existing research or study fair dealing provisions, or may be covered by statutory licence if done for educational purposes.⁷⁷

8.54 In particular, publishers argued that the market for data and text mining is still developing, and that solutions to the perceived problem have not had a chance to evolve. For example, John Wiley & Sons submitted that:

There is currently little or no uniform understanding of what TDM (text/data mining) actually is, nor how best it can be enabled or supported. From our experience, there is little consistency across TDM projects as far as activities, processes and results are

72 House of Commons Business, Innovation and Skills Committee, *The Hargreaves Review of Intellectual Property: Where next?* (2012), 19.

73 UK Government, *Modernising Copyright: A Modern, Robust and Flexible Framework* (2012), 37.

74 Ibid.

75 This analysis was supported in submissions from the ADA and ALCC, *Submission 213* and R Xavier, *Submission 146*.

76 Copyright Agency/Viscopy, *Submission 249*; APRA/AMCOS, *Submission 247*; Australian Directors Guild, *Submission 226*; Australian Copyright Council, *Submission 219*.

77 Copyright Agency/Viscopy, *Submission 249*; Australian Publishers Association, *Submission 225*.

concerned, let alone definitions around content access methods and protocols or standard licensing terms.⁷⁸

8.55 The Association of Learned and Professional Society Publishers (ALPSP) argued that ‘publishers are not blocking access to articles for text and data mining—publishers are reporting that current requests are very low, and in the main, they are granted’.⁷⁹ Therefore, it was suggested that solutions lie in co-operation between users and publishers to create licensing solutions.⁸⁰ Exceptions, it was argued, would not create an environment conducive to collaboration:

Data and text mining solutions are best found in market-based initiatives, like proactive voluntary licensing, that offer faster and more flexible ways to adapt to changing market needs and preferences. These solutions must be based on collaboration between users and publishers. Value proposals and business models for publishers in the field of data and text mining are only now emerging, and publishers are experimenting with various contractual and operational models.⁸¹

8.56 Publishers also argued that licensing helps offset publishers’ costs to support content mining on a large scale, and that increases in costs ‘could act as a significant disincentive to publishers to continue to invest in programmes to enrich and enhance published content, which in turn facilitates greater usage and encouragement’.⁸²

8.57 Publishers warned that ‘the relative immaturity of the TDM market should not be considered as indicative of market failure demanding legislative intervention’.⁸³

8.58 Other stakeholders were concerned about the reach of any data and text mining exception into commercial operations.⁸⁴ For example, Telstra recognised the value of data and text mining ‘in the context of research, education and culture’, but was opposed to reform that would allow the use of data mining tools or software for commercial exploitation. For example:

an offshore data-miner that scrapes (or copies) data from an online Australian database, such as a telephone directory. The data-miner then uses the scraped content to establish a competing business, without the need to source, verify, supplement or format the content. The data-miner also avoids the need to employ Australian staff, or to invest in the creation or development of content.⁸⁵

78 John Wiley & Sons, *Submission 239*.

79 ALPSP, *Submission 199*.

80 Australian Publishers Association, *Submission 225*.

81 IASTMP, *Submission 200*.

82 John Wiley & Sons, *Submission 239*. The APA argued that cost implications arise because ‘crawling can affect platform performance and response times, and may require the development and maintenance of parallel content delivery systems; costs are then incurred to ensure that adequate performance and access (whether for licensed or unlicensed users) is maintained: Australian Publishers Association, *Submission 225*.

83 John Wiley & Sons, *Submission 239*; Australian Publishers Association, *Submission 225*; ALPSP, *Submission 199*.

84 Telstra Corporation Limited, *Submission 222*; Australian Broadcasting Corporation, *Submission 210*; Cyberspace Law and Policy Centre, *Submission 201*. The Cyberspace Law and Policy Centre stressed that ‘there is a need to manage access to address, technical, competitive and commercial risks’.

85 Telstra Corporation Limited, *Submission 222*.

8.59 IASTMP argued that publishers are increasingly providing licensing solutions for commercial text mining and that they should be allowed to continue providing or facilitating customised data and text mining solutions.⁸⁶

Facilitating research and study

8.60 A number of stakeholders argued that data and text mining should be permitted, drawing on the principle of ‘non-expressive’ use, or uses that do not trade on the underlying or expressive purpose of the work.⁸⁷

8.61 For example, the Australian Industry Information Association argued that it is important for legislative reform to encourage research, development and competition in the data analytics field. It suggested a specific exception to allow data and text mining for the purposes of ‘comparison, classification or analysis’ would not negatively impact on the original data provider’s rights and commercial interests because the technology is not intended to reprint the original data, but to provide a synthesised result. These outcomes do not interfere with the economic value of the copyright material nor compete with it.⁸⁸

8.62 Similarly, others referred to use of academic materials and journals that could be considered as ‘transformative’ uses.⁸⁹ The ADA and ALCC suggested that data and text mining, as a subset of transformative use may be best supported by a flexible, open ended exception:

uses which may have been characterised as transformative, such as text and data mining, but may be better seen as ‘non-expressive’ or ‘orthogonal’ uses. Fair use in the US provides the flexibility for new technologies to develop which may straddle the two definitions, and similarly providing courts with the tools to deem when such uses will unreasonably harm the copyright owner.⁹⁰

8.63 A number of submissions referred to the importance of data and text mining for non-commercial research and study.⁹¹ However, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) argued that the commercial/non-commercial distinction is not useful, since:

such a limitation would seem to mean that ‘commercial research’ must duplicate effort and would be at odds with a goal of making information (as opposed to illegal copies of journal articles, for example) efficiently available to researchers ... As noted, much research is conducted through international collaboration. If the laws in Australia are more restrictive than elsewhere or if the administration of any rights system is cumbersome or onerous and creates excessive cost for research, then that might be expected to impact on the desirability of Australia as a research destination.⁹²

86 IASTMP, *Submission 200*.

87 ADA and ALCC, *Submission 213*; Australian Industry Group, *Submission 179*.

88 AIIA, *Submission 211*. See also Internet Industry Association, *Submission 253* who also supported an exception around copying for the purposes of extracting information.

89 ADA and ALCC, *Submission 213*; R Xavier, *Submission 146*; M Rimmer, *Submission 138*.

90 ADA and ALCC, *Submission 213*.

91 CSIRO, *Submission 242*; Telstra Corporation Limited, *Submission 222*; M Rimmer, *Submission 138*.

92 CSIRO, *Submission 242*.

Non-consumptive uses and fair use

8.64 Australian copyright law should recognise that the reproduction of copyright material is a necessary part of the effective functioning of technology in the digital environment. The fact that copyright material has been copied—for example by a search engine—should not, of itself, provide a full answer to the question of infringement. Copyright exists to protect the expression of ideas and facts, rather than the facts themselves.

8.65 The ALRC proposes that the fair use exception proposed in Chapter 4 should be applied when determining whether uses of copyright material for the purposes of caching, indexing or text mining infringes copyright.

8.66 The ALRC also proposes that ‘non-consumptive’ uses should be one of the illustrative purposes of the fair use exception.⁹³ This should signal that uses that fall within the broader category of ‘non-consumptive use’ are more likely to be fair than uses that do not fall into this, or any other, category of illustrative purpose. However, this does not mean that all non-consumptive uses will be fair. A wider inquiry into the fairness factors is necessary and crucial.

8.67 There is a spectrum along which uses of copyright material may, to a greater or lesser extent, be said to be ‘trading on’ the underlying and expressive purpose of the copyright material. In the ALRC’s view, the fairness factors, including the nature and use of the copyright material; the portion of the material that is taken; and the impact on the potential market for the work provides a flexible framework for balancing the interests of users and rights holders in a way that specific exceptions cannot.

8.68 In the event that fair use is not enacted, the ALRC proposes an alternative, namely, fair dealing for ‘non-consumptive’ uses. This fair dealing exception would require consideration of whether the use is fair, having regard to the same fairness factors that would be considered under the general fair use exception.

8.69 The *Copyright Act* should define ‘non-consumptive’ use to mean uses of copyright material ‘that do not trade on the underlying or expressive purpose the material’. The ALRC proposes that ‘non-consumptive use’ be defined, because unlike the existing fair dealing provisions—such as parody and satire or reporting the news—it is not immediately clear what this term means.

8.70 The section below explains the ALRC’s reasoning as to why caching and indexing, and data and text mining, should be considered under fair use.

Caching and indexing

8.71 In the ALRC’s view, the use of copyright material for caching, indexing and other similar functions that are necessary in the digital environment should not infringe copyright. Indeed, the fact that no company has been sued in Australia for caching and

93 See Ch 4.

indexing might suggest that rights holders consider such activities to be ‘fair use’, or that such uses do not sufficiently prejudice existing markets to warrant litigation.⁹⁴

8.72 There are strong arguments that lack of protection for such activities comparable to other jurisdictions may create an environment of uncertainty which could have an impact on investment decisions about whether to operate in Australia or contribute to increases in the cost of providing services to the Australian public, such as internet streaming of television programs. The development of cloud computing services will also increase the need for temporary copies to be made.

8.73 Further, it appears difficult to draft a specific exception for caching and indexing that would be technology neutral, and that would accommodate the different interests of the parties. Technology reliant on copying will continue to evolve, and the *Copyright Act* needs to be to adaptive to such technological changes.

8.74 In the ALRC’s view, a general fair use exception may provide more flexibility to consider the impact on the market than blanket exception that permits caching and indexing.

8.75 The caching and communication of content located behind a ‘paywall’ or ‘subscription content’ is unlikely to be fair use.⁹⁵ On the other hand, as argued by Pandora, where a licence has been obtained to communicate recordings and temporary copies are made for the purposes of exercising that licence, this should not be subject of further licensing.⁹⁶ This appears on its face to be a non-consumptive use more likely to be fair (that is, incidental copying in order to exercise a right).

8.76 The *Copyright Act* contains a number of disparate provisions that deal with ‘temporary copying’ that are intended to cover different forms of caching or copying that is required as part of the way a technology functions. It is undesirable to have multiple provisions that do not adequately cover the full spectrum of caching activities. If fair use is enacted, these existing exceptions should be repealed.

Data and text mining

8.77 For similar reasons, the ALRC considers that the use of copyright material for data and text mining should also be considered under the fair use exception, in determining whether copyright is infringed.

8.78 There is not enough evidence of market failure to warrant a specific exception to deal with data and text mining, and the benefits of the data analytics industry are capable of being maximised through collaboration between researchers and publishers. In particular, the ALRC considers that voluntary licensing should be pursued for commercial uses of data and text mining.

94 In consultations the ALRC heard that rights holders consider some caching or indexing activities as having an implied or zero licence, rather than accepting that the use is fair.

95 Having regard to the portion that is copied and the fact that there is market for subscription access to copyright material.

96 Pandora Media Inc, *Submission 104*.

8.79 In the ALRC's view, fair use would not undermine emerging market solutions for data analytics. Rather, the availability of licensing solutions would be one factor in determining whether a data or text mining use is fair. The fairness factors are intended to provide a framework within which a number of competing interests can be balanced. In respect of data and text mining, these can include but are not limited to:

- the amount of copyright material that was copied;
- whether the data or text mining will be used for a non-commercial purpose;
- whether the use is to facilitate education and research;
- the existence of any agreed industry guidelines; and
- whether the copying resulted in an end use that is considered transformative and that does not trade on the underlying expressive purpose of the copyright material that is copied.⁹⁷

Proposal 8-1 The fair use exception should be applied when determining whether uses of copyright material for the purposes of caching, indexing or data and text mining infringes copyright. 'Non-consumptive use' should be an illustrative purpose in the fair use exception.

Proposal 8-2 If fair use is enacted, the following exceptions in the *Copyright Act* should be repealed:

- (a) s 43A—temporary reproductions made in the course of communication;
- (b) s 111A—temporary copying made in the course of communication;
- (c) s 43B—temporary reproductions of works as part of a technical process of use;
- (d) s 111B—temporary copying of subject-matter as a part of a technical process of use; and
- (e) s 200AAA—proxy web caching by educational institutions.

Proposal 8-3 If fair use is not enacted, the *Copyright Act* should be amended to provide a new fair dealing exception for 'non-consumptive' use. This should also require the fairness factors to be considered. The *Copyright Act* should define a 'non-consumptive' use as a use of copyright material that does not directly trade on the underlying creative and expressive purpose of the material.

97 Whether a use is transformative can be considered when applying fair use. See Ch 10.

