

11. Incidental or Technical Use and Data and Text Mining

Contents

| | |
|--|-----|
| Summary | 249 |
| Incidental or technical use | 250 |
| Current exceptions | 252 |
| Assisting innovation and lawful consumption | 254 |
| Fair use | 255 |
| An illustrative purpose | 257 |
| Fair dealing for incidental or technical use | 259 |
| Specific exception for incidental or technical use | 259 |
| Safe harbour scheme | 260 |
| Data and text mining | 260 |
| Non-expressive use | 261 |
| Current law | 262 |
| Fair use | 264 |

Summary

11.1 Incidental or technical uses are essential to the operation of the internet, networks, and other technologies that facilitate lawful access to copyright material. The ALRC concludes that current exceptions in the *Copyright Act* are uncertain and do not provide adequate protection for such uses. Such uncertainty has adverse effects on innovation, incentives to build new services and products, and Australia's competitiveness as a place for technological investment. The current exceptions should be repealed.

11.2 The ALRC recommends that 'incidental or technical use' be an illustrative purpose of fair use. A flexible fair use exception will better accommodate technological change and foster lawful innovation and use of copyright material. If fair use is not enacted, a new fair dealing exception should be introduced, and this should include 'incidental or technical use' as a prescribed purpose.

11.3 Data and text mining refers to technologies that are used to analyse copyright material for patterns, trends and other useful information. The ALRC concludes that the fair use exception should be used to determine whether data and text mining constitute copyright infringement.

Incidental or technical use

11.4 Copyright owners have an exclusive right to reproduce and communicate their work to the public.¹ Without a licence or an exception, reproduction and communication to the public of copyright material constitute infringement.

11.5 This can be problematic in a digital environment where material ‘cannot be handled without copying it’.² Reproduction and communication of copyright material has not only become ubiquitous, but necessary for the effective and efficient functioning of the internet, networks, and technological processes that facilitate lawful consumption of copyright material. For example, a reproduction and communication is required every time a person watches a DVD, reads a webpage,³ or streams a video from the internet.⁴ In contrast, no such reproduction or communication is required in a non-digital context when a person reads a book or a magazine.

11.6 One example of incidental or technical use discussed in this Inquiry related to caching and indexing by search engines. 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. 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.⁵

11.7 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.⁶

11.8 Search engines, web hosts and other internet intermediaries rely on indexing and caching for their efficient operation.⁷ Other parties also rely on caching and indexing to facilitate streaming services and to improve the speed of database searches.⁸

1 See *Copyright Act 1968* (Cth) s 31. Further, the definition of ‘material form’ in s 10 suggests that electronic reproduction of copyright material will constitute copyright infringement.

2 iiNet Limited, *Submission 186*.

3 Internet browsers store ‘cached’ copies of a webpage to enable immediate retrieval when a person revisits the same page. Caching can also 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. Thus, 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.

4 Temporary cache and ‘buffer’ copies are usually made in the course of streaming content from the internet to ensure seamless experience for the user. For example, the 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: ABC, *Submission 210*.

5 Google Guide, *Cached Pages* <www.googleguide.com/cached_pages.html> at 30 July 2012.

6 Ibid. A website can specifically prevent a crawler from accessing parts of their website that would otherwise be publically viewable, by inserting a piece of code called ‘robot.txt’.

7 See eg, iiNet Limited, *Submission 186*.

8 Screenrights, *Submission 215*; ADA and ALCC, *Submission 213*.

11.9 Stakeholders also emphasised the importance of the cloud computing sector to the future of Australia's digital economy.⁹ Innovative cloud-based services rely heavily on technological processes that involve incidental or technical reproduction and communication of copyright material.¹⁰

11.10 Ericsson argued that the extra copying required in the digital environment results in businesses requiring more 'time and money to acquire necessary permissions', which are not required in relation to analog works.¹¹ The impact of such concerns was highlighted by the experience of Pandora, an internet streaming radio service that struggled to negotiate licences it needed to operate in Australia. Pandora argued that it needs to make permanent copies to deliver its services to the Australian public, and that it should not be required

to separately negotiate licences to make copies of recordings where it secures a licence to communicate the recordings and the copies are made purely for the purposes of exercising that licence.¹²

11.11 The policy question for this Inquiry was whether unlicensed incidental or technical uses should infringe copyright and, if so, under what circumstances. Can the copyright system facilitate the efficient operation of digital technologies to promote innovation and ensure wide access to copyright material, while acknowledging and respecting authorship and creation?

11.12 There has been growing international consensus that certain unlicensed incidental or technical reproduction should not be viewed as infringing. For example, Maria Pallante, Director of the US Copyright Office, has observed that

new technologies have made it increasingly apparent that not all reproductions are the same. Some copies are merely incidental to an intended primary use of a work, including where primary uses are licensed, and these incidental copies should not be treated as infringing.¹³

11.13 This sentiment was echoed by a number of stakeholders in this Inquiry, including the ACCC:

In the digital environment there has been an increase in the use and copying of copyright material, in ways that appear to be quite incidental to the production of the primary good or service being produced. For example, copyright material copied by internet intermediaries for caching purposes. Similarly, some use and copying of copyright material may currently be 'unauthorised' in circumstances where this use has little, if any, detrimental impact on incentives for copyright creation.¹⁴

9 See, eg, OzHub, *Submission 148*; CCH Australia Ltd, *Submission 105*; K Bowrey, *Submission 94*; eBay, *Submission 93*.

10 Telstra Corporation Limited, *Submission 222*; Australian Industry Group, *Submission 179*.

11 Ericsson, *Submission 151*.

12 Pandora Media Inc, *Submission 104*.

13 M Pallante, 'The Next Great Copyright Act' (2013) 36(3) *Columbia Journal of Law & the Arts* 315, 325. See also eBay, *Submission 93*.

14 ACCC, *Submission 165*. See also K Bowrey, *Submission 554*; eBay, *Submission 93*.

11.14 The Hargreaves Review also recommended that the UK Government push to build into the EU framework an exception allowing uses of a work enabled by technology which do not trade on the underlying 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 a technology works ... This is not about overriding the aim of copyright—these uses do not compete with the normal exploitation of the work itself—indeed, they may facilitate it.¹⁵

Current exceptions

11.15 The *Copyright Act* contains a number of exceptions that deal with temporary reproductions. These include:

- ss 43A and 111A—allowing for the temporary reproduction of a work and an adaptation of a work or an audiovisual item as part of the ‘technical process of making or receiving a communication’;¹⁶
- ss 43B and 111B—providing that copyright is not infringed by a temporary reproduction ‘incidentally made as a necessary part of a technical process’ of using a copy of the work or subject matter;¹⁷
- s 116AB—allowing 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 to facilitate efficient access to that material by that user or other users;¹⁸
- s 200AAA—allowing automated caching by computers operated by or on behalf of an educational institution;¹⁹ and
- ss 47, 70 and 107, allowing copying to make broadcasts technically easier and to enable the making of repeat or subsequent broadcasts.²⁰

11.16 Stakeholders suggested that these exceptions were not adequate to deal with caching and indexing and other technical or incidental uses in the digital environment. For example, ss 43A and 111A only permit a ‘temporary’ reproduction, but copyright

¹⁵ I Hargreaves, *Digital Opportunity: A Review of Intellectual Property and Growth* (2011), 47.

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 has been suggested that ss 43B and 111B could apply to caching by search engines: 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. Reforms to the safe harbour provisions are outside this Inquiry’s Terms of Reference.

¹⁹ Such caching can only be done ‘in response to action by the users’ and ‘the reproductions and copies are made by the system merely to facilitate efficient later access to the works and other subject-matter by users of the system’.

²⁰ These ‘ephemeral’ copying provisions are discussed in Ch 19.

material may need to be stored in a cache for long periods of time.²¹ Similarly, it was submitted that the exceptions allow only a single reproduction, whereas the digital environment demands and that multiple reproductions are necessary.²²

11.17 Stakeholders also submitted that exceptions do not adequately protect certain technical or incidental communications of copyright material, such as when a search engine communicates search results to a user.²³ The Copyright Advisory Group—Schools observed that there was merit in considering whether exceptions for ‘temporary communications’ are required in the same way as for ‘temporary reproductions’:

For example, the upload of a work to a Learning Management System would involve a reproduction of that work, but the display of that work in class (via connection to a laptop and/or interactive whiteboard) or accessing the content by a student or staff from the cloud or a centralised content repository, may also result in one or more electronic transmissions comprised in the right of communication to the public when the content is transmitted from the Learning Management System (LMS) to a laptop, monitor or electronic whiteboard.²⁴

11.18 Optus argued that s 43A only allows copies to be made after the user requests a download of the material and that this was not consistent with what happens on a practical level, where ‘a copy is created in a cache in anticipation of download by other users’.²⁵ Similarly, it was suggested that it is unclear whether s 200AAA could facilitate ‘active’ forms of caching, whereby a school selects what material needs to be cached.²⁶

11.19 Burrell and others suggested that the limitation in ss 43A(2) and 111A(2), that the copy not be an infringing copy, makes the exception unworkable for caching of any significance:

Any entity that sets up their system to cache all (or all popular) communications is likely, at some point, to capture copies from both infringing and non-infringing communications without any knowledge on their part.²⁷

21 Google, *Submission 217*; Optus, *Submission 183*.

22 See, eg, ABC, *Submission 210*, suggesting that the ‘communication of streamed program content may also be more effectively managed by service providers and intermediaries caching content at various points in the technical delivery chain, such as through the use of edge servers and mirror sites within content delivery systems, rather than streaming content from centralised servers and data warehouses’. See also Telstra Corporation Limited, *Submission 222*.

23 ACCC, *Submission 165*. See also Telstra Corporation Limited, *Submission 222*; Law Council of Australia, *Submission 263*; Google, *Submission 217*; iiNet Limited, *Submission 186*.

24 Copyright Advisory Group—Schools, *Submission 231*.

25 Optus, *Submission 183*. Similar concerns were expressed by the Law Council of Australia who argued that the terminology of ss 43A and 111A casts some doubt over the scope of the provision, for example, whether it covers proxy caching intended to facilitate access to users other than the ones involved in ‘a communication’: Law Council of Australia, *Submission 263*.

26 Law Council of Australia, *Submission 263*.

27 R Burrell, M Handler, E Hudson, and K Weatherall, *Submission 278*. See also K Weatherall, *Internet Intermediaries and Copyright: An Australian Agenda for Reform* (2011), Policy Paper prepared for the Australian Digital Alliance.

11.20 The concerns from stakeholders were summed up by the Law Council of Australia, which stated that the legal position in relation to caching in Australia is ‘confusing, overlapping, incoherent and in some cases redundant’ 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’.²⁸

Assisting innovation and lawful consumption

11.21 Stakeholders submitted that the uncertainty around caching and indexing inhibits, or has the potential to inhibit, innovation and investment in cloud computing and other products and services that rely on such incidental or technical uses.²⁹ While copyright holders suggested that nothing in the *Copyright Act* has impeded search engines from providing services to Australians, it remains the case that the ‘servers that these services run on are all located overseas, and mostly in the US, because they simply can’t operate in Australia’.³⁰

11.22 Some stakeholders were concerned to ensure that Australia’s regulatory framework puts it on the same footing as other jurisdictions.³¹ For example, the Australian Industry Group emphasised the need for Australia’s regulatory framework to be consistent and competitive with other jurisdictions:

A copyright framework that prohibits critical or routine activities related to the digital economy that are permitted in other markets may discourage domestic innovation or lead to commercial or research activities staying or moving offshore.³²

11.23 Burrell and others argued that, on principle, incidental or technical uses should be excluded from infringement. They suggested that rights holders should not be able to ‘double dip’ or otherwise expand the reproduction and communication rights to demand licence fees for each individual copy made to facilitate lawful uses of copyright material.³³ If reproductions that are necessary and ubiquitous risk infringement, this may have the effect of increasing transaction costs (as more licences are required) and stifling the creation of innovative services in the digital economy.³⁴ The ACCC said that

Transaction costs can also arise as a result of uncertainty regarding whether certain incidental uses currently breach the *Copyright Act*, as resolving uncertainties can be expensive and in some instances, require litigation. These uncertainties could cause

28 Law Council of Australia, *Submission 263*.

29 See, eg, Internet Industry Association, *Submission 744*; Google, *Submission 600*; AIMIA Digital Policy Group, *Submission 261*; R Giblin, *Submission 251*; Optus, *Submission 183*; Australian Industry Group, *Submission 179*.

30 R Xavier, *Submission 816*, noting that ‘Australia’s lack of fair use does not mean that works by Australians are not subject to fair use on the internet ... It just means that none of these services can be provided from Australia, and that the revenue from running them flows overseas’. See also Optus, *Submission 183*.

31 AIMIA Digital Policy Group, *Submission 261*; EFA, *Submission 258*; ADA and ALCC, *Submission 213*; Optus, *Submission 183*; Australian Industry Group, *Submission 179*; R Xavier, *Submission 146*.

32 Australian Industry Group, *Submission 179*.

33 R Burrell, M Handler, E Hudson, and K Weatherall, *Submission 716*.

34 Ericsson, *Submission 151*.

third parties to avoid developing products that involve beneficial incidental copying, which can in turn limit product innovation and competition.³⁵

Fair use

11.24 A number of stakeholders—including technology companies that rely heavily on caching and indexing—suggested that uncertainty could be removed by replacing the current exceptions with fair use.³⁶ Some compared the uncertain situation in Australia to the US, where caching and indexing is done in reliance on a ‘well established fair use doctrine that permits this activity’.³⁷

11.25 Stakeholders also suggested that fair use is more suitable than specific exceptions, because it can accommodate uses that ‘don’t exist yet, or haven’t yet been foreseen’.³⁸ For example, Google submitted that closed exceptions

are antithetical to how the internet works and the dynamic nature of the creativity enabled by the internet. Australia’s system of closed-purpose, prescriptively described exceptions means that new and innovative uses of copyright materials that do not fall within the technical confines of an existing exception are not capable of being permitted by exceptions, no matter how creative the new use, or how strong the public interest in enabling that new use may be.³⁹

11.26 Others suggested that it would be difficult to draft purpose-based exceptions for caching, indexing and other internet functions without ‘some technology specificity’.⁴⁰ Telstra argued that redrafting ‘based on today’s technical knowledge and standards is likely to render the exception obsolete in the context of future innovations’.⁴¹

11.27 In Chapter 4, the ALRC makes the case that fair use is flexible, and can accommodate for technological change in ways that specific exceptions cannot. Australian copyright law should recognise that the reproduction and communication of copyright material is a necessary part of the effective functioning of technology in the digital environment. The fact that copyright material has been reproduced or communicated—for example by a search engine—should not, of itself, infringe copyright. The question should be answered by an analysis of whether such uses are fair.

11.28 New and unforeseen technical or incidental uses—beyond caching and indexing—will arise in the digital environment. The ALRC considers that fair use is sufficiently flexible to determine whether such uses should be permitted, based on an assessment of fairness. Importantly, fair use requires market considerations to be taken into account and this should protect the interests of copyright owners. A specific or blanket exception may not adequately provide for such protection.

35 ACCC, *Submission 165*. See also Google, *Submission 217*.

36 R Burrell, M Handler, E Hudson, and K Weatherall, *Submission 278*; Yahoo!7, *Submission 276*; EFA, *Submission 258*; Telstra Corporation Limited, *Submission 222*; Google, *Submission 217*; ADA and ALCC, *Submission 213*; Law Institute of Victoria, *Submission 198*; iiNet Limited, *Submission 186*.

37 Yahoo!7, *Submission 276*; iiNet Limited, *Submission 186*.

38 Telstra Corporation Limited, *Submission 222*.

39 Google, *Submission 600*.

40 Telstra Corporation Limited, *Submission 222*; ADA and ALCC, *Submission 213*.

41 Telstra Corporation Limited, *Submission 222*.

11.29 In the ALRC's view, caching and indexing that are uncertain under the current exceptions would likely be fair, under the fair use exception recommended in this Report. However, whether a use is fair, must in each instance, be assessed after considering the following fairness factors.

The purpose and character of the use

11.30 Whether a use is 'transformative' will be a key question in applying the Australian fair use exception.⁴² This requires an examination of the extent to which a new work merely 'supersedes' or 'supplants' the original work or whether the new work is 'for a different expressive purpose from that for which the original was created'.⁴³ A number of US court decisions have held caching and indexing to be transformative.

11.31 In *Field v Google*, a US court found that copies held in Google's cache were 'transformative' because they allowed users to:

- access content when the page was inaccessible;
- detect changes made to a page;
- understand why the page was responsive to their original query.⁴⁴

11.32 Similarly, in *Perfect 10, Inc v Amazon.com, Inc*, thumbnails of artistic works that were communicated by Google's cache were considered to be 'highly transformative', because a search engine provides an entirely new use for the original work, turning the image from a use of artistic expression into an 'electronic reference tool'.⁴⁵

11.33 The ALRC considers that caching and indexing are transformative and that this would weigh heavily in favour of fair use. Other incidental or technical uses may not be transformative, but may nevertheless be fair for other reasons.

11.34 Whether an incidental or technical use is commercial will also be relevant. A commercial purpose will tend to weigh against a finding of fair use,⁴⁶ but this will not always be the case.⁴⁷ For example, the US Copyright Office suggests that buffer copies made in the course of internet music streaming is fair use, despite being done for a commercial activity. Buffer copies are not a 'superseding use that supplants the original'. Rather, they are necessary and 'non-exploitative' and the purpose 'is to enable a use that has been authorised by the copyright owner and for which the copyright owner typically has been compensated'.⁴⁸ It considered that the commercial aspect in such cases 'can best be described as of minimal significance'.⁴⁹

42 See Ch 5.

43 N Weinstock Netanel, 'Making Sense of Fair Use' (2011) 15 *Lewis and Clark Law Review* 715, 768. See also *Campbell v Acuff-Rose Music Inc* (1994) 510 US 569, 579.

44 *Field v Google Inc* (2006) 412 FSupp 2d 1106 (District Court of Nevada), 1119.

45 *Perfect 10, Inc v Amazon.com, Inc*, 508 F 3d 1146 (9th Cir, 2007), 15468.

46 *Harper & Row Publishers, Inc v Nation Enterprises* (1985) 471 US 539, 585.

47 *Blanch v Koons*, 467 F 3d 244 (2nd Cir, 2006).

48 US Copyright Office and the Library of Congress, *DMCA Section 104 Report* (2001), 133.

49 *Ibid.*

The nature of the copyright material used

11.35 The nature of the copyright material will be a relevant consideration. For example, reproduction of published material is more likely to be fair use than in the case of unpublished material.⁵⁰

The amount and substantiality of the part used

11.36 This factor considers how much of work is taken, and how important was that taking in the context of the plaintiff's work. One question is whether the incidental or technical use takes only what is 'reasonably necessary' for a particular technical function.⁵¹

11.37 In some instances, they may require a small portion of the work to be reproduced, while in others, a whole of a work may need to be produced. For example, in *Kelly v Arriba Soft Corporation*, the court recognised that an internet search engine needs to engage in wholesale copying in order to provide any meaningful responses to search queries.⁵² While holding that Arriba's use of thumbnail images to be fair use, it held the search engine's display of full size images was not fair use because this was not transformative and resulted in 'substantial adverse effects to the potential market for Kelly's original works'.⁵³

Effect of the use upon the market

11.38 The effect of the use on the market will be a significant factor, and may often depend on how transformative a use is. A use that is transformative is less likely to substitute for the original work, and therefore less likely to cause harm to the market.

11.39 For the market harm factor to dictate against fair use, the harm to the market should be substantial, rather than minor or remote. Findings about whether incidental or technical uses have 'independent economic significance' may be relevant in determining whether such uses affect the market of the original work. For example, the US Copyright Office suggests that buffer copies made in the course of streaming have no 'economic value independent of the performance it enables' which can harm the market. Rather, such copying merely 'facilitates an already existing market for the authorised and lawful streaming of works'.⁵⁴

An illustrative purpose

11.40 Incidental or technical uses appear to be a good example of fair use in the digital environment. In particular, the first and fourth factors, when applied to technical and incidental uses—such as caching and indexing or network functions—that facilitate lawful access to copyright material will tend to weigh heavily in favour of fair use.

50 *Campbell v Acuff-Rose Music Inc* (1994) 510 US 569, 586.

51 *Ibid*, 588.

52 *Kelly v Arriba Soft Corporation*, 280 F 3d 934 (9th Cir, 2002).

53 *Ibid*, 948.

54 US Copyright Office and the Library of Congress, *DMCA Section 104 Report* (2001), 139.

11.41 The ALRC recommends that the fair use exception include an illustrative purpose for ‘incidental or technical use’. This does not mean that all incidental or technical uses are fair, but it will signal that such uses are more likely to be fair than non-technical or non-incidental uses. However, when determining whether a particular technical or incidental use is fair, the fairness factors should all be considered.

11.42 In the Discussion Paper, the ALRC proposed that ‘non-consumptive use’—being use of copyright material that does not trade on the underlying and expressive purpose of the copyright material—be an illustrative purpose in the fair use provision.⁵⁵ The ALRC considered that this purpose could cover technical or incidental uses as well as non-expressive uses, such as data and text mining (discussed below).

11.43 Stakeholders submitted that ‘non-consumptive’ as an illustrative purpose was vague and uncertain, and required further elaboration.⁵⁶ It was submitted that:

- it is difficult to ascertain what ‘trade on’ or ‘underlying expressive purpose’ means;⁵⁷
- the term may rule out completely any use that is in some way ‘consumptive’ or commercial;⁵⁸
- the term does not exist in any other international legislation;⁵⁹ and
- there is no ‘bright line’ as to when, how or to what extent, infringing use will not trade on the underlying or expressive purpose.⁶⁰

11.44 Some stakeholders were not convinced that the phrase ‘non-consumptive’ could neatly cover both technical and incidental uses and those that are not merely facilitative but are non-expressive in nature. They suggested that the conflation of the two uses under the umbrella term ‘non-consumptive use’ may cause difficulties in the application of fair use.⁶¹

11.45 After further consideration, the ALRC has decided to recommend the narrower and precise ‘incidental or technical use’ as an illustrative purpose.

55 Australian Law Reform Commission, *Copyright and the Digital Economy*, Discussion Paper 79 (2013), Proposal 8–1.

56 CSIRO, *Submission 774*; News Corp Australia, *Submission 746*; Australian Film/TV Bodies, *Submission 739*; AFL, *Submission 717*; Cricket Australia, *Submission 700*; Springer Science and Business Media, *Submission 639*.

57 Intellectual Property Committee, Law Council of Australia, *Submission 765*; Queensland Law Society, *Submission 644*; Cyberspace Law and Policy Centre, *Submission 640*; R Xavier, *Submission 531*.

58 R Burrell, M Handler, E Hudson, and K Weatherall, *Submission 716*. They suggest that ‘non-consumptive’ use may rule out reproductions that are made as part of lawful consumption activities, such as reading a book or viewing copyright material in private. See also Music Council of Australia, *Submission 647*.

59 News Corp Australia, *Submission 746*.

60 Australian Film/TV Bodies, *Submission 739*.

61 R Burrell, M Handler, E Hudson, and K Weatherall, *Submission 716*.

Fair dealing for incidental or technical use

11.46 The ALRC also recommends that, if fair use is not enacted, a new fair dealing exception should be introduced.⁶² This new fair dealing exception should include ‘incidental or technical use’ as a prescribed purpose.

11.47 The 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 fair use exception. Applying the fair use or fair dealing exceptions to incidental or technical uses should produce the same result.

Specific exception for incidental or technical use

11.48 If neither fair use nor the new fair dealing exception is enacted, the ALRC suggests that the existing exceptions for temporary reproductions could be repealed in any event, and replaced with a new specific exception for incidental or technical use.

11.49 In the ALRC’s view, this is a less flexible solution to fair use or new fair dealing. Nonetheless, a new specific exception may be warranted to alleviate uncertainty surrounding the current exceptions.

11.50 If a new specific exception is introduced, it should be technology neutral and should fill in the gaps that exist under the current exceptions. This includes recognition that incidental or technical uses require both reproduction and communication of copyright material. The references to ‘temporary’ should be removed. It should also be made clear that copyright is not infringed by using copyright material for caching and indexing by search engines, where that facilitates a lawful use.

11.51 This specific exception could also be confined to incidental or technical uses that facilitate lawful use of copyright material. This would prevent the use of the exception to aid infringing practices. Such reform would provide greater clarity and certainty to innovators that services or products relying on technical or reproduction to facilitate the lawful use of copyright material, will not infringe copyright.

11.52 In considering such a new specific exception, the Australian Government may wish to consider specific exceptions found in other jurisdictions. For example, in Canada, the *Copyright Act 1985* (Can) provides that it is not infringement to make a reproduction of a work that forms part of an essential process, for the purpose of facilitating a use that is not an infringement to copyright, and the reproduction exists only for the duration of the technical process.⁶³

11.53 Canada also has an exception for ‘network services’. A person who provides services related to the internet or another digital network do not, solely by providing

62 See Ch 6.

63 *Copyright Act 1985* (Can), s 30.71.

those means, infringe copyright in the telecommunication or reproduction of a work or other subject matter.⁶⁴

11.54 The Copyright Review Committee (Ireland) has recommended an exception that would allow reproductions which are temporary, transient or incidental and that have no independent economic significance. The reproduction must be an essential and integral part of a technological process with the sole purpose of enabling either a transmission in a network between third parties by an intermediary or a lawful use.⁶⁵ This mirrors a European Union Information Society Directive, which provides that states should include a mandatory exception to the right of reproduction in respect of certain temporary acts of reproduction.⁶⁶ The Directive also leaves open to member states to provide for other non-mandatory exceptions to the right of communication.⁶⁷

Safe harbour scheme

11.55 The exceptions discussed in this chapter do not remove the need for a safe harbour scheme. Australia's safe harbour scheme limits the remedies available against 'carriage service providers' for copyright infringement that takes place on their systems, which they do not control, initiate or direct. The scheme is now being reviewed, and is outside the ALRC's Terms of Reference.⁶⁸

11.56 However, it should be noted that even if fair use or the new fair dealing exception for incidental or technical use were enacted, internet intermediaries and others may still need to rely on a safe harbour scheme in other circumstances.

Recommendation 11-1 The exceptions for temporary uses and proxy web caching in ss 43A, 111A, 43B, 111B and 200AAA of the *Copyright Act* should be repealed. The fair use or new fair dealing exception should be applied when determining whether incidental or technical uses infringes copyright.

Data and text mining

11.57 Data and text mining has been defined as automated analytical techniques that work by 'copying existing electronic information, for instance articles in scientific

64 Ibid s 31.1. This is subject to a number of conditions including that the person: does not modify the material other than for technical reasons; ensures that directions in a manner consistent with industry practice are followed; 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 of the work or other subject matter: s 31.3.

65 Copyright Review Committee (Ireland), Department of Jobs, Enterprise and Innovation, *Modernising Copyright* (2013), 116.

66 *Directive 2001/29/EC of the European Parliament and of the Council on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society*, (entered into force on 22 June 2001), art 5.

67 Ibid, art 5(4).

68 See Australian Government Attorney-General's Department, *Revising the Scope of the Copyright 'Safe Harbour Scheme'*, Consultation Paper (2011), 5. The safe harbour provisions are found in div 2AA of the *Copyright Act*.

journals and other works, and analysing the data they contain for patterns, trends and other useful information'.⁶⁹

11.58 Data and text mining is becoming increasingly important in a number of research sectors, including medicine, business, marketing, academic publishing and genomics.⁷⁰ Employing technology to mine journal databases has been referred to as 'non-consumptive' research, because it does not involve human reading or viewing of the works.⁷¹ 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.⁷²

11.59 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.⁷³ 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.⁷⁴

Non-expressive use

11.60 There has been growing recognition that data and text mining should not be infringement because it is a 'non-expressive' use. Non-expressive use leans on the fundamental principle that copyright law protects the expression of ideas and information and not the information or data itself. For example, consider a computer algorithm employed to search through a text to obtain metadata, which discovers two facts about Moby Dick:

first, that the word 'whale' appears 1119 times; second, that the word 'dinosaur' appears 0 times. While a whale is certainly central to the expression contained in

⁶⁹ UK Government Intellectual Property Office, *Consultation on Copyright* (2011), 80.

⁷⁰ 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.

⁷¹ 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).

⁷² See, eg, UK Government, *Consultation on Copyright: Summary of Responses* (2012), 17.

⁷³ 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).

⁷⁴ Cyberspace Law and Policy Centre, *Submission 201*.

Moby Dick, this data is not. Rather, metadata of this sort ... is factual and non-expressive, and incapable of infringing the rights of copyright holders.⁷⁵

11.61 Academics use this example to argue that ‘acts of copying that do not communicate the author’s original expression to the public do not generally constitute copyright infringement’.⁷⁶ They suggest that to the extent that data and text mining do not substitute for the author’s original expression, such non-expressive uses

are properly considered equivalent to (or a subset of) highly transformative uses: their ‘purpose and character’ is such that they do not merely supersede the objects of the original creation.⁷⁷

11.62 Similarly, Burrell and others submitted that uses that treat copyright material as mere data—rather than for its expressive value—do not compete with the original works and should not be treated as falling within the scope of the copyright owner’s rights.⁷⁸

11.63 Similar thinking was evidenced in the Hargreaves Review, which recommended an exception for uses of works enabled by technology which do not trade on the underlying and expressive purpose of the work. As a result of the recommendation, the UK Government will introduce an exception that allows a person who already has access to a work (whether under license or otherwise) to copy the work as part of a technological process of analysis and synthesis of the content of the work for non-commercial purposes.⁷⁹

Current law

11.64 There is no exception in the *Copyright Act* that covers data and text mining. Where the data or text mining processes involve the copying, digitisation, or reformatting of copyright material without permission, it may give rise to copyright infringement.

11.65 One issue is whether data and text mining, if done for the purposes of ‘research or study’, would be covered by the fair dealing exception. 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.⁸⁰ Nor is it clear whether copying for text mining would fall under the s43B exception relating to temporary reproduction of works as part of a technical process, but it seems unlikely.

75 M Jockers, M Sag and J Schultz, *Brief of Digital Humanities and Law Scholars as Amici Curiae in Authors Guild v. Hathitrust* (2013), 18.

76 *Ibid.*, 1609.

77 M Jockers, M Sag and J Schultz, *Brief of Digital Humanities and Law Scholars as Amici Curiae in Authors Guild v. Hathitrust* (2013).

78 R Burrell, M Handler, E Hudson, and K Weatherall, *Submission 716*.

79 Intellectual Property Office, *Data Analysis for Non-commercial Research* (2013).

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

11.66 A number of stakeholders argued that data and text mining should be covered by fair use,⁸¹ drawing on the principle of non-expressive use, or uses that do not trade on the underlying or expressive purpose of the work.⁸² Others suggested that data and text mining are properly considered as ‘transformative’ uses.⁸³

11.67 The Australian Industry Information Association argued that it is important for legislative reform to encourage research, development and competition in the data analytics field.⁸⁴ Universities Australia suggested that subjecting data and text mining to fair use would put Australian universities

on a level playing field with their counterparts in the US (who rely on fair use to engage in non-consumptive uses such as data mining and text mining for socially useful purposes) as well as the UK (who will soon have the benefit of a stand-alone exception for non-commercial data mining and text mining).⁸⁵

11.68 The Commonwealth Scientific and Industrial Research Organisation (CSIRO) agreed that if laws in Australia are more restrictive than elsewhere, the increased cost of research would make Australia a less attractive research destination.⁸⁶

11.69 A number of stakeholders suggested that data and text mining should be limited to non-commercial research and study.⁸⁷ However, the 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.⁸⁸

11.70 Other stakeholders agreed.⁸⁹ Google submitted that there are clear public benefits to facilitating data and text mining ‘regardless of whether this occurs within the confines of a university or other public research institution, or in the private sector’.⁹⁰

81 Internet Industry Association, *Submission 253*; Google, *Submission 217*; Society of University Lawyers, *Submission 158*; R Xavier, *Submission 146*.

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

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

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

85 Universities Australia, *Submission 754*.

86 CSIRO, *Submission 242*.

87 AFL, *Submission 717*; Cricket Australia, *Submission 700*; CSIRO, *Submission 242*; Telstra Corporation Limited, *Submission 222*; M Rimmer, *Submission 138*.

88 CSIRO, *Submission 242*. The problematic distinction between commercial/non-commercial was also highlighted by Cyberspace Law and Policy Centre, *Submission 640* and John Wiley & Sons, *Submission 239*.

89 Universities Australia, *Submission 754*; Google, *Submission 600*; Cyberspace Law and Policy Centre, *Submission 201*.

90 Google, *Submission 600*.

11.71 On the other hand, publishers opposed an exception for data and text mining and suggested that ‘the relative immaturity of the text/data mining market should not be considered as indicative of market failure demanding legislative intervention’.⁹¹

11.72 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’.⁹² Therefore, it was suggested that solutions lie in cooperation between users and publishers to create licensing solutions.⁹³ 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 ... 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.⁹⁴

11.73 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’.⁹⁵

Fair use

11.74 The ALRC considers that the unlicensed use of copyright material for non-expressive purposes, such as data and text mining, should be considered under the fair use exception recommended in this Report.

11.75 The ALRC agrees that non-expressive use can be considered a subset of transformative use. To the same extent that transformative use is not an illustrative purpose, the ALRC does not consider it necessary to include ‘non-expressive use’ or ‘data and text mining’ in the list of illustrative purposes.

11.76 Arguments in favour of considering data and text mining under a fair use exception, rather than introducing a new specific exception, largely parallel the more general arguments for introducing fair use. Data and text mining can ‘cover a range of activities which do or may not raise the same issues’.⁹⁶ It is clear that data and text mining technologies are still evolving and they will become useful across a wide range

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

92 ALPSP, *Submission 199*.

93 Australian Publishers Association, *Submission 225*.

94 IASTMP, *Submission 200*.

95 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*.

96 Intellectual Property Committee, Law Council of Australia, *Submission 765*. See also John Wiley & Sons, *Submission 239* which submitted that ‘there is currently little or no uniform understanding of what TDM actually is, nor how best it can be enabled or supported’.

of sectors in the economy, both commercial and non-commercial. The ALRC considers that fair use is sufficiently flexible to balance the competing interests between ‘copyright owners on the one side and academic and commercial users of data mining techniques on the other’.⁹⁷

11.77 Whether a use is fair must, in each instance, be assessed after considering the following fairness factors.

The purpose and character of the use

11.78 Data and text mining for illustrative purposes of fair use, such as ‘research or study’, ‘education’, ‘library or archive use’, are more likely to be fair. For example, the ALRC considers that the illustrative purpose of ‘research and study’ under fair use would allow data and text mining on the same grounds as the exception being implemented in the UK. This broadly aligns with the view of publishers, who had little problems with data mining for non-commercial purposes where a person has subscribed to the content that is being mined.⁹⁸

11.79 A finding that data and text mining is transformative would weigh heavily in favour of fair use. For example, to the extent that data and text mining allows ‘for the creation of new information, new aesthetics, new insight and understanding’,⁹⁹ its use may be considered transformative.

11.80 Data and text mining for a commercial purpose would generally disfavour a finding fair use, but not always. The Cyberspace Law and Policy Centre submitted that data mining may be done in relation to commercial medical research, and it is not clear that the commerciality ought always to be decisive, when all the fairness factors are considered.¹⁰⁰

The nature of the copyright material used

11.81 Copyright exists to protect the expression of ideas and facts, rather than the facts themselves. US courts have held that the scope of fair use is greater with respect to factual than non-factual works.¹⁰¹ In addition, it has also been held that ‘the second factor may be of limited usefulness where the creative work of art is being used for a transformative purpose’.¹⁰²

The amount and substantiality of the part used

11.82 The amount and substantiality needed will depend on the purpose and character of the use. The ALRC envisages that many data and text mining exercises, to be useful, will involve reproduction of entire works. Fair use case law in the US makes it clear

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

98 International Association of Scientific Technical and Medical Publishers, *Submission 560*.

99 P Leval, ‘Toward a Fair Use Standard’ (1989–1990) 103 *Harvard Law Review* 1105, 1111.

100 Cyberspace Law and Policy Centre, *Submission 640*.

101 *Basic Books, Inc. v. Kinko’s Graphics Corp.*, 758 F Supp 1522 (SNDY, 1991), 1533.

102 *Bill Graham Archives v. Dorling Kindersley, Ltd.*, 448 F3d 605 (2nd Cir, 2006), 612.

that reproduction of a whole of a work can, depending on the circumstances, amount to fair use.¹⁰³

Effect of the use upon the market

11.83 The effect on the market would be a relevant factor. Where the use is non-expressive or highly transformative, there will be good arguments that such uses are not a substitute for the original work, and therefore cannot directly harm the market for the original. For the market factor to work against fair use, the unlicensed use must harm ‘traditional, reasonable, or likely to be developed’ markets.¹⁰⁴

11.84 The ALRC appreciates the arguments that licensing solutions are being developed for data and text mining. However, the mere availability of a licence should not mandate that unlicensed uses are never fair. However, where a licence is offered on reasonable terms, it will be more difficult to argue that the unlicensed use is fair. This will go against a finding of fair use, especially where the use is also commercial and non-transformative.

103 *The Authors Guild Inc v HathiTrust*, WL 4808939 (SDNY, 2012).

104 *Princeton University Press v Michigan Document Services, Inc*, 99 F 3d 1381 (6th Cir, 1996), [26].