Ownership and control over publicly accessible platform data

Teresa Scassa
Faculty of Law, University of Ottawa, Ottawa, Canada

Received 12 February 2018
Revised 9 July 2018
10 December 2018
10 January 2019
Accepted 14 January 2019

Abstract

Purpose – The purpose of this paper is to examine how claims to "ownership" are asserted over publicly accessible platform data and critically assess the nature and scope of rights to reuse these data.
Design/methodology/approach – Using Airbnb as a case study, this paper examines the data ecosystem that arises around publicly accessible platform data. It analyzes current statute and case law in order to understand the state of the law around the scraping of such data.
Findings – This paper demonstrates that there is considerable uncertainty about the practice of data scraping, and that there are risks in allowing the law to evolve in the context of battles between business competitors without a consideration of the broader public interest in data scraping. It argues for a data ecosystem approach that can keep the public dimension issues more squarely within the frame when data scraping is judicially considered.
Practical implications – The nature of some sharing economy platforms requires that a large subset of their data be publicly accessible. These data can be used to understand how platform companies operate, to assess their compliance with laws and regulations and to evaluate their social and economic impacts. They can also be used in different kinds of data analytics. Such data are therefore sought after by civil society organizations, researchers, entrepreneurs and regulators. This paper considers who has a right to control access to and use of these data, and addresses current uncertainties in how the law will apply to scraping activities, and builds an argument for a consideration of the public interest in data scraping.
Originality/value – The issue of ownership/control over publicly accessible information is of growing importance; this paper offers a framework for approaching these legal questions.

Keywords Copyright, Sharing economy, Platform economy, Data ownership, Data scraping

Paper type Research paper

Introduction

Internet platform companies host a significant amount of data on their sites. Some, although not all, of this information is personal information and often much of it is user contributed. While the data – and the ability of the public to access and view them – are an important part of the business model of these platforms, these data are often of significant interest to those beyond the immediate sphere of platform users. In a data-driven economy, publicly accessible data can be an important resource. Because of this, questions arise as to who has the right to control access to and use of such data, and in what ways and circumstances such control can be exercised. The answers to these questions will have important implications for innovators, researchers, civil society and even governments.

This paper considers questions around ownership and control of publicly accessible data, using Airbnb as a case study. Airbnb is a major, global platform that hosts a great quantity and variety of data. The nature of its business requires that these data are made publicly accessible – in order for hosts to share information about their rental units and for potential or actual guests to be able to browse this information, and add to it with comments and reviews. Airbnb is an interesting case study because its operations have had significant impacts on many cities, raising questions about, among other things, the platform's effects on the cost and availability of long-term accommodation, its impact on incumbent short-term accommodation providers, the incidence of discrimination in Airbnb rentals and pricing and the extent to which the platform is used to support full-scale commercial ventures. The data hosted on the

The author gratefully acknowledges the support of the Social Sciences and Humanities Research Council of Canada. Thanks to Nathan Hoo and Joël Boisvert for their research assistance.

Online Information Review
Vol. 43 No. 6, 2019
pp. 986-1002

1468-4527
DOI 10.1108/OIR-02-2018-0053

Airbnb site can be scraped and analyzed so as to provide important insights into these issues. In the absence of adequate voluntary data sharing by a company, data scraping remains a primary source of such data. Airbnb platform data are also of interest to a wide range of businesses, many of which are not in direct competition with Airbnb, but which instead offer services – including analytics – related to Airbnb's activities.

This paper considers the complex "ecosystem" of users of Airbnb platform data in order to identify and assess the nature and extent of public and private interests in using such data. In order to understand the impacts of Airbnb – and in order to build innovative or opportunistic businesses that draw upon Airbnb's data – it is necessary to be able to access, harvest and manipulate these data. While any company may protect its confidential commercial information from undesired access and reuse, the rights of companies to protect and limit the reuse of publicly accessible data are more tenuous. Any such rights may also be subject to countervailing users' rights. The legal status of publicly available information is therefore important. In a rapidly evolving data economy, legal uncertainties in relation to ownership of and rights of access to data risk being resolved by litigation between business competitors, which risks overlooking and unduly limiting the strong public interest in access to and use of such data. This paper argues that a data ecosystem approach to publicly accessible platform data is necessary to prevent the normative framework for data scraping from being unduly shaped by the platforms themselves.

Literature review

Although a great deal has been written about the sharing economy and its impacts, relatively little scholarly attention has been paid to the massive quantities of data that are amassed by platform companies, and more specifically about issues relating to access to and reuse of these data. This has started to change somewhat, as developments in the artificial intelligence (AI) sector raise questions about access to data to train algorithms (Geiger *et al.*, 2018). The data-sharing practices of social media companies have also come under recent scrutiny with the Cambridge Analytica scandal, with preoccupations in this area relating to the automated extraction and reuse of personal information (Information Commissioner's Office, 2018). Much less has been written about access to personal and non-personal platform data for other types of research.

Access to data about platform companies' activities can be difficult to obtain. For example, the impacts of short-term rental platforms on the availability and affordability of long-term accommodation in cities has led to numerous studies and reports, many of which lament the lack of easily accessible data (Jamasi and Hennessy, 2016; Sawatzky, 2015; Vancouver City Council, 2016; City of Toronto, 2016; Cutler, 2015). As documented by Scassa (2017), a number of cities have resorted to using data scraped either by civil society actors or by consultants. Scraped data have formed the basis for a number of studies and report about short-term rental platform economy impacts (Scassa, 2017; Sawatzky, 2015; Clampet, 2014).

The practice and the legality of data scraping have received relatively little academic attention, although this is beginning to change as such activities become increasingly widespread and more commercially significant (Snell and Care, 2013). Quite apart from the issue of whether scraping infringes copyright by taking a substantial part of an original selection or arrangement of data, legal scholars have considered whether data scraping is a form of trespass to chattels (Din, 2015; Warner, 2002), or whether (in the USA) it violates the *Computer Fraud and Abuse Act* (Din, 2015; Hirschey, 2014).

Legal disputes over data scraping have begun to heat up, although the bulk of this case law involves competing businesses. Some cases involve data scraping by relatively direct competitors (*Century 21 Canada Limited Partnership* v. *Rogers Communications Inc.*, 2011; *Trader Corporation* v. *CarGurus, Inc.*, 2017; *Ryanair Ltd* v. *PR Aviation BV*, 2015), others

involve companies in the same field but operating with a different business model. For example, the news aggregator Meltwater has generated litigation over its scraping of news headlines on both sides of the Atlantic (*Newspaper Licensing Agency Ltd and Ors* v. *Meltwater Holding BV and Ors*, 2011; *Associated Press* v. *Meltwater U.S. Holdings, Inc.*, 2013). In other cases, data scrapers harvest data for new forms of analytics. In the USA, litigation between LinkedIn and a number of companies that actively scrape its data promises to produce interesting case law around the legality and the limits of such practices (Conger, 2016). A very recent US case, *Sandvig* v. *Sessions* (2018) challenges the application of the *Computer Fraud and Abuse Act* to data-scraping activities carried out for research purposes. However, the emerging case law will not provide clear answers to all questions regarding the legitimacy of data scraping as a means of acquiring publicly accessible data. In the first place, some of the decisions turn on the form in which data are presented. For example, if photographs are scraped (*Century 21 Canada Limited Partnership* v. *Rogers Communications Inc*., 2011; *Trader Corporation* v *CarGurus, Inc.*, 2017), or newspaper headlines (*Associated Press* v. *Meltwater U.S. Holdings, Inc.*, 2013; *Newspaper Licensing Agency Ltd and Ors* v. *Meltwater Holding BV and Ors*, 2011), these forms may attract a greater level of copyright protection than would be available for compiled data. Second, most, though not all, data-scraping disputes involve commercial competitors and such disputes are less likely to raise issues of users' rights or the public interest. This means not only that these issues are rarely explored in the emerging court decisions, but that the decisions themselves may shape the law in ways that develop robust concepts of the rights of data "owners" while doing nothing to consider or articulate the nature and scope of users' rights. There are a couple of notable exceptions in the USA (*Sandvig* v. *Sessions*, 2018; *hiQ Labs, Inc.* v. *LinkedIn Corporation*, 2017) but these cases largely address US-specific legislation that, depending on its interpretation, might criminalize data scraping in some circumstances.

The thirst for data in the AI sector has prompted considerable legal interest around the world in creating text and data-mining exceptions to copyright infringement (Geiger *et al.*, 2018). The data-scraping discussion thus now also takes place in the shadow of broader concerns over rights to access and reuse data for AI research and processes. These discussions have also touched on other issues relevant in the data-scraping context, including the impact of terms of service (TOS) that prohibit automated extraction of data, and the use of technological protection measures (TPMs).

Methods

The analysis of data scraping was carried out using a single platform company – Airbnb – as a case study in order to provide a rich context for the assessment of data scraping. The users of Airbnb data are as diverse as the users to which the data are put, which helps ease out the complex competing interests.

The research behind this paper involved a detailed study of Airbnb's platform, including a web-based review to identify users of Airbnb data, how data are accessed by these users and the uses to which the data are put. Studies, reports and other documents that relied upon scraped Airbnb data were analyzed, and materials regarding difficulties in accessing Airbnb data, and critiques of voluntary releases of data by Airbnb were gathered from a review of print and online sources.

Because data scraping is the primary means by which Airbnb's platform data are harvested for reuse, the research also involved a scan of the primary (laws, case law) and secondary (scholarly articles and commentary) legal literature regarding data scraping. Legal materials from Canada, the USA and the UK were analyzed to provide the basis for a discussion of the law of data scraping. Since data scraping is also impacted by aspects of private law, the legal documents that structure Airbnb's operations were reviewed.

These include its TOS, privacy and other policies. Legal claims in these documents, including ownership or control of data, contractual limitations such as prohibited conduct with respect to data and site usage, and privacy commitments were considered, along with their interaction with applicable laws.

Results

Airbnb impacts

According to its website, Airbnb operates in over 191 countries and over 65,000 cities worldwide (Airbnb). It also claims to have more than 3m listings, with over 150m travelers finding accommodation through its platform. Like many platform companies, Airbnb exploits the rhetoric of facilitating small-scale sharing of individuals' surplus resources. It claims to help ordinary individuals monetize their underutilized living space, whether in the form of a spare room, or their entire unit while on vacation. However, with the rise in the popularity of the site, concerns have grown over the use of the platform by those who make entire units available year-round; as well as hosts with multiple available units. Indeed, a growing concern is that long-term accommodation is being converted by urban entrepreneurs into short-term rental accommodation via Airbnb (Office of the Attorney General of the State of New York, 2014). Such activities have raised serious concerns about the impact of the platform company, particularly in cities where there is a shortage of long-term accommodation as well as problems with affordable housing. In addition, policy makers and civil society groups have argued that Airbnb contributes to gentrification (Cox, 2017; Office of the Attorney General of the State of New York, 2014; Jamasi and Hennessy, 2016); undermines or changes the character of neighborhoods (Scassa, 2017); and creates nuisances (Scassa, 2017). Condominium associations and landlords have also raised concerns over the impact of short-term rentals on the character of their buildings, the over-exploitation of shared spaces, disregard of noise, non-smoking and other policies and security (Scassa, 2017). Incumbent short-term rental industries have also complained that Airbnb hosts are not subject to the same level of regulation and do not pay taxes, thus enabling them to unfairly compete within the market (Reyes, 2015).

Airbnb data

Not all data collected by Airbnb are exposed on its platform, much data – including personal information of its hosts and guests – is kept confidential. Airbnb generates its own confidential analytics based upon its data. Nevertheless, the company makes a range of different data publicly accessible on its platform. In this respect, Airbnb is different from other platform companies such as Uber. In order to meet its goal of connecting hosts with guests, the Airbnb platform must necessarily display information about the nature and size of available units, their price, general location and general availability. The site also contains photographs of units, verbal descriptions of units, their amenities and their location, information about hosts, reviews (which may contain information about guests and what they did during their stay, information about host and guest interactions; information about the unit, about the dates of the stay and about the area in which the unit is located). All of these data come from multiple sources, including hosts (who provide photos, verbal descriptions, availability information, etc.); guests (who provide reviews); and Airbnb (which provides templates for information as well as in some cases photographs). Although the primary purpose of these data is to provide information about specific units that are for rent, taken together and subject to analytics, they can provide rich information about the number of rental units available in particular areas, the availability of those units over time, their price (varying by time of year and over time) and much more. These data can be used in analytics to assist those seeking to enter the market to appropriately price their rental units. This information can also be used to understand

the extent to which units are really just "excess space" or are diverted from the pool of available long-term accommodation. They can provide valuable insight into a range of other issues relating to long- and short-term accommodation, tourism, tax avoidance, zoning and by-law infractions, breaches of leases and other contracts. Public facing Airbnb data have also been used to explore issues of discrimination in the provision of short-term rental accommodation (Wang *et al.*, 2015). The uses of the data are limited only by the creativity of the users.

Airbnb data ecosystem

The "ecosystem" metaphor has become widely used in scientific and technological literature. It builds upon the concept of natural ecosystems as dynamic, evolving contexts in which there are multiple interdependencies (Harrison *et al.*, 2012). Zuiderwijk *et al.* (2014, p. 20) refer to metaphorical ecosystems as reflecting "the dynamic interaction between different factors in an area." More simply put, according to Nardi and O'Day (1999, p. 49), an ecosystem is defined as "a system of people, practices, values, and technologies in a particular local environment." Harrison *et al.* (2012, p. 906) observe that: "The dynamic of ecosystems is one of flow and movement – people, ideas, activities, and tools in motion as the ecosystem evolves continuously in the form of components that 'adjust and are adjusted in relation to each other, always attempting and never quite achieving a perfect fit.'"

The ecosystem metaphor for publicly accessible Airbnb data provides a way to understand the context that is framed in terms of interdependencies and interrelationships rather than one in which there is merely a sequence of one-on-one relationships between a company and the various legitimate and possibly illegitimate users of its data. Airbnb data are contributed and created by different actors; they are used by different actors for different purposes; and they are capable of describing or of contributing to descriptions of phenomena and experiences within real communities. This concept of Airbnb data as part of an ecosystem therefore shifts the paradigm from one of corporate ownership/control of data in the context of a company's distinct relationships with different users to one in which there are a network of different claims to rights and or interests in the data.

A thriving data ecosystem has arisen around Airbnb's publicly accessible data. A variety of users may access and use the data in different ways and using different means. One of these means is that for which the site was created – prospective guests may browse the site in order to find information about short-term accommodation in the cities they plan to visit. Hosts may also browse the site to compare their unit with others in the same area, and in order to gain information about how they might price their unit or better present it to the public. These uses are explicitly permitted in the TOS of the site.

Civil society organizations also make use of Airbnb data. This is often done to raise awareness of issues regarding the availability and affordability of long-term accommodation. In the case of Airbnb, there are some high-profile examples. Activist Tom Slee, for example, has scraped Airbnb data and hosts a website that makes these data as well as analysis and studies based on the data publicly available (www.tomslee.net). Slee has also, in the past, made the code he uses for scraping data publicly available. Another high-profile Airbnb activist, Murray Cox, operates the website InsideAirbnb.com. Through his site he provides commentary and analysis of scraped Airbnb data.

Journalists and researchers are also users of Airbnb platform data. Typically this is scraped data. The journalists or researchers may scrape the data themselves (e.g. Sawatzky, 2016; Wang *et al.*, 2015) or they may use data scraped by others (such as Slee or Cox, above) (e.g. Majoribanks, 2016). Not all Airbnb data that are used in research are scraped. For example, Edelman *et al.* (2017) who studied discrimination over the Airbnb platform gathered their data through contact with hosts over the platform.

Similar to researchers, journalists pursuing Airbnb-related stories may scrape their own data, or they may rely upon data already scraped by others (e.g. Clampet, 2014; Said, 2015).

Airbnb's publicly accessible data ecosystem also includes a range of different opportunistic businesses. These are businesses that have sprung up around Airbnb and that depend – to greater or lesser extents – on the continued operation of Airbnb. They also depend on ongoing access to Airbnb data. An example is the company Airdna. Airdna (airdna.co) offers its clients a variety of data analytics services. These include "market reports and other data products that feature occupancy rates, seasonal demand, and revenue generated by short-term rentals" (airdna.co). According to Airdna.co, its reports and analytics "are based on Airbnb data gathered from information publicly available on the Airbnb website" (www.airdna.co/methodology). Airdna is also linked to another business, Rentingyourplace.com, which offers consulting services to prospective Airbnb hosts. Airdna is not the only analytics company to mine Airbnb data. Other companies include, but are not limited to, Beyond Pricing (beyondpricing.com), SmartHost (smarthost.co.uk), Everbooked (www.everbooked.com) and PriceLabs (www.pricelabs.co).

Not all businesses that rely on Airbnb data offer analytics for those who seek to participate in the short-term rental market. Because of the importance of Airbnb data to urban planners, researchers and governments (among others), consulting companies may scrape Airbnb data in order to provide a broader range of consulting services. For example, Host Compliance LLC (2016) produced a report based on scraped data for the City of Vancouver. Airdna, mentioned above, also provides broader consulting services based on Airbnb publicly accessible data (Stulberg, 2016).

Another category of businesses makes use of publicly available Airbnb data in order to provide a different kind of service. There are a growing number of detective agencies – either general practices or ones specifically focused on short-term rental detection – that use data on the Airbnb platform in order to determine whether units in their clients' buildings are being illegally rented through the platform. While some agencies may use manual techniques (personally searching through listings) others are using automated search tools to crawl through short-term rental listings (www.buildingsnitch.com).

As can be seen from the above, many opportunistic businesses are not competitors of Airbnb in a strict sense, although in some cases they may compete indirectly or in sub-markets for Airbnb data. In the case of Airbnb, a company might scrape Airbnb data in order to combine it with other available data to provide information to those considering offering a unit for rent on a platform such as Airbnb. This information might include recommendations as to price point, peak rental periods and so on. This does not necessarily compete with Airbnb – in fact, it might complement its business by making it easier for people to list properties on the platform. However, should Airbnb choose to provide similar analytics or to sell access to its data for these purposes, then the scraping activities arguably undermine these activities. Whether this is characterized as fair or as unfair competition may turn on whether Airbnb is seen as entitled to control its public facing data as an intellectual property asset. Competition may be difficult to define or identify in a rapidly evolving context in which new applications are constantly being discovered for data, and in which the platform company's control over its publicly accessible data would give it the ability to control the commercial exploitation of these applications. Within this data ecosystem, therefore, there is some uncertainty as to the boundaries of legitimate and illegitimate conduct. These in turn are tied to notions of what rights exist in publicly accessible platform data – including rights to control and exclude and rights to access and use.

The boundaries between legitimate and unfair competition are in part at the root of the current litigation between LinkedIn and the opportunistic companies that scrape its data; these companies have found markets for data and/or analytics based on LinkedIn data and

LinkedIn objects to their commercial exploitation of these markets (Conger, 2016). While not competing directly with LinkedIn's primary business (hosting a business networking site), these other activities arguably exploit for profit the data LinkedIn has collected. If LinkedIn has rights to control its publicly accessible data, then this exploitation by others of the data is a breach of those rights. Yet these activities might just as easily be characterized as innovation using publicly accessible data.

Discussion

Platform companies have a number of legal tools that they can use in efforts to assert control over the publicly accessible data hosted on their sites. These tools can be divided into three categories. The first category involves legal claims based upon "ownership" rights – whether this involves ownership of intellectual or personal property. The second involves limitations on access that are ultimately supported by law. These limitations may involve contractual terms or technological barriers that limit access. The third involves privacy rights.

Ownership

Ownership claims asserted by platform companies are of two kinds: intellectual property rights and rights in chattels. Intellectual property rights associated with data involve copyright, and, in the EU, may also involve database rights. Chattel (personal property) rights are asserted in relation to the physical infrastructure that hosts the data.

Intellectual property rights. A platform company's rights in its publicly accessible data can be complex, particularly where a significant quantity of that data is user contributed. This complexity is reflected in the Airbnb TOS. This document distinguishes between content contributed by its members ("Member Content") and content that Airbnb itself makes available over the site ("Airbnb Content"). A third category of content – "Collective Content" – reflects the combination of both Member and Airbnb content. Airbnb asserts copyright in its own content, but does not claim copyright in Member Content, asserting only a perpetual, non-exclusive worldwide license to use and disseminate it. Nevertheless, as the host and compiler of the "Collective Content," Airbnb may have a copyright in the overall compilation on its site, as discussed below. It is certainly possible for a party to have a copyright in a compilation even if different parts of the overall compilation are contributed by others who retain copyright in their respective contributions. Article 5.2 of the Airbnb TOS states that: "The Airbnb Platform, Airbnb Content, and Member Content may in its entirety or in part be protected by copyright, trademark, and/or other laws of the United States and other countries." Article 5.3 prohibits certain uses of the Collective Content that are consistent with claims of copyright in such content, and Article 5.4 provides a limited license to "access and view" the Collective Content. These provisions are consistent with a claim to copyright in the compilation that is the Collective Content.

Copyright claims depend upon the existence of a "work" in which copyright subsists. Protected works must fall into one of the categories of literary, artistic, dramatic or musical works. They must also be "original." Some web-based data are represented in forms or ways that independently constitute works. For example, a photograph is an artistic work; photographs are also a way in which data can be represented. Text is a literary work, but can also be a representation of data. In some cases, data scraping from websites has involved the scraping of photographs (*Trader Corporation* v. *CarGurus, Inc.*, 2017; *Century 21 Canada Limited Partnership* v. *Rogers Communications Inc*., 2011) or text (*Associated Press* v. *Meltwater U.S. Holdings, Inc*., 2013; *Century 21 Canada Limited Partnership* v. *Rogers Communications Inc*., 2011). In such cases, it can be argued that the scrapers have violated copyright in those works by reproducing them without permission.

In other cases, however, scrapers have merely extracted data (*Ryanair Ltd* v. *PR Aviation BV*, 2015). In general terms, copyright law does not protect "facts," which are considered to be in the public domain – free for anyone to use (Tamaroff, 2011). However, a compilation consisting of an original selection or arrangement of facts may be protected (Newell, 2011; Hugenholtz, 2017). Some authors explore what it takes to have an original selection or arrangement of facts (Newell, 2011; Leaffer, 2007), with most concluding that the threshold for protection of such a compilation is quite low. Nevertheless, while compilations of fact may be relatively easily protected as "works" the extent of protection is considered "thin," as infringement requires a substantial taking of either the original selection or arrangement. The taking of public domain facts themselves is not, on its own, infringement. Thus, any claim to copyright infringement with respect to the data hosted on a platform site will depend upon whether the site hosts a compilation of data that are original by virtue of their selection or arrangement, and whether the scraper has extracted a substantial part of that original selection or arrangement. Arguably, a platform company, by deciding what information users must provide and by creating the formats by which it is arranged online, has produced an original selection and arrangement of data. However, to the extent that all short-term rental platforms require the same categories of information about rental units and their availability, a court might consider that the selection of data is merely routine and not original. And, while the arrangement of these data may vary from platform to platform, providing a sufficient degree of originality, a data scraper who extracts the data from this context and stores it according to his or her own protocols may not be taking a substantial part of the platform's original arrangement. Thus, copyright arguments against data scraping (as opposed to the scraping of text or of photographs) are complex and contingent.

Copyright protection for compilations of facts is roughly equivalent in the EU. However, in 1996 the EU also passed a Database Directive which established a *sui generis* regime for the protection of databases. This directive creates a right in the "maker" of a database who has made a substantial investment in the creation of the database (Tamaroff, 2011). The database right makes it an infringement to extract or reuse some or all of the contents of the database. While this seems to provide more extensive protection than copyright, recent court cases in Europe have considerably restricted the application of the Directive, and some now question its usefulness in protecting many compilations of data (Hugenholtz, 2017; Newell, 2011). The significance of the database right in protecting against data scraping is therefore in doubt. This may explain why debate began to stir within the EU over whether a new "data ownership" right should be created (Hugenholtz, 2017). Other data ownership/access issues that are emerging in the Big Data context include whether text and data-mining activities infringe copyright in the works that are used (which can include compilations of data) or whether and in what circumstances these activities might constitute fair dealing/fair use (Geiger *et al.*, 2018; Sobel, 2017). These disputes highlight the significance of public interest exceptions to copyright principles and the challenges of delineating the boundaries of ownership rights.

Rights to use/access. Just as there are some legal arguments that can be asserted by platform companies to protect their publicly accessible data from data scraping, legal arguments are also available to those who seek to scrape and reuse such data.

Copyright law creates a balance between the rights of owners of copyright and those of users; this balance serves the greater public interest in the broad dissemination of works and in the free and open dissemination of ideas. If it can be successfully argued that there is copyright in the compilation of data hosted on a platform website, a scraper might be able to argue that their actions in taking a substantial part of the selection or arrangement of the data constitute fair use (in the USA) or fair dealing. Fair use/fair dealing rights particularly

(though not exclusively) favor non-commercial uses and ones that support activities protected by freedom of expression values such as research, criticism or comment. In the USA, the creation of new or "transformative" works can also be fair use.

The strength and scope of such a defense may vary from one jurisdiction to another. Fair use in the USA, for example, is a more expansive defense than fair dealing under Canadian law. The purpose for which the data are scraped may have some bearing on the success of such a defense – scraping for research purposes may be more likely to be considered fair than scraping to establish a business – particularly one that is in full or partial competition with the target of the scraping. Similarly, non-commercial uses may be considered fairer than commercial ones – although not all commercial uses will be unfair.

The debates over whether text and data-mining activities infringe copyright law provide an interesting comparison with the issues raised by data scraping. The thirst for data to feed data analytics and to develop machine learning has led to a need in some sectors to absorb data from copyright-protected sources. To do so, entire texts must be scanned or entire compilations of data absorbed. This wholesale copying would infringe the owners' copyrights unless it falls within an exception to infringement. In the USA, the flexible fair use exception is considered by some to justify text and data mining (e.g. Cox, 2015). In the EU, by contrast, a specific exception is thought necessary, and the scope and wording of such an exception is currently a matter of debate (Geiger *et al.*, 2018). The challenge is to properly assess the different interests at play – including the public interest – and to strike the appropriate balance. This is so whether a solution is achieved by interpreting existing laws or making new ones. Similar challenges exist with the scraping of publicly accessible data, although these are made more complex by the contingent and uncertain nature of any copyright in a compilation of data from the outset.

Several factors may influence fair dealing analyses when it comes to data-scraping activities. For example, if a website's terms of use prohibit data scraping, the fact that a user must breach contractual obligations in order to harvest the data might mitigate against a finding that this is fair use or fair dealing. Some scholars have argued that rights holders should not be allowed to alter fair dealing/fair use exceptions through contracts of adhesion (Elmahjub and Suzor, 2017; DiValentino, 2014; Elkin-Koren, 1997) but this is an area where the case law is still unsettled. The presence of robots.txt protocols to signal that scraping is not permitted might also be considered relevant in a fair use/fair dealing analysis. The automated, repeated and large-scale nature of some scraping practices might also be considered presumptively unfair.

In its litigation with LinkedIn, data scraper hiQ asserted freedom of speech rights under California law in support of an argument that it was entitled to collect and use publicly available information. The court was not persuaded by this argument (*hiQ Labs, Inc.* v. *LinkedIn Corporation*, 2017). Furthermore, in most instances, freedom of speech rights are constitutional guarantees and relate to obligations owed by the government; they are not applicable between private parties. Nevertheless, the argument touches on the tension between ownership rights and the freedom of ideas, knowledge and information. Where data are publicly accessible, what should be the boundaries of private rights to control and limit the reuse of that information, and what role should the law play in reinforcing those boundaries?

Airbnb appears to have been relatively restrained when it comes to addressing the scraping of its platform data. There are currently no records of lawsuits initiated against Airbnb data scrapers. In cases where Airbnb data have been scraped in order to produce reports or studies on the impact of the platform in cities, the company has asserted that scraped data are unreliable and unfit for purpose (Hiltzik, 2015; Sawatzky, 2015).

Although the scope and subsistence of copyright in any compilation of data is uncertain and contingent, the availability of fair dealing/fair use defenses is also uncertain. Legal uncertainty

of this kind may be enough on its own to deter those with limited resources from engaging in contested actions. Researchers, civil society organizations – others without deep pockets – are generally not able to contest cease-and-desist letters. This may explain why the data-scraping cases that have reached the courts have almost exclusively involved corporations. A rare exception, *Sandvig* v. *Sessions* (2018), involved claims by researchers; this litigation was supported by the American Civil Liberties Union.

Chattel rights. In some data-scraping cases in the USA, plaintiffs have argued that the data scraper is engaging in a trespass to chattels. Trespass to chattels is a tort action that is available when personal property is interfered with, either by intentionally dispossessing another of a chattel or by using or interfering with a chattel that is in someone else's possession (Warner, 2002). The "chattels" in a data-scraping case are the servers on which the data are stored; the interference is remote and electronic. Essentially, the argument is that the data scraper, by using "crawlers" and "robots" to scrape data from the server, has substantially interfered with the chattel, and is therefore liable. These arguments have had mixed success, and are more likely to prevail where the scraping activities are so frequent or so extensive that they overburden and impede a server's ability to function properly. Thus, where data-scraping activities cause a server to crash or even to respond more slowly to legitimate requests for data, it is more likely that the plaintiff will succeed with a tort claim in trespass to chattels. Nevertheless, even scraping activities that have little or no discernable impact on the host server may be actionable as trespass to chattels (Din, 2015).

Contractual or technological restrictions

Platform companies regularly use TOS to set the rules of conduct for their sites. These are a form of contract that bind the user of the site either through their express consent (clicking an icon that indicates that they accept the terms of use) or through their conduct (continuing to browse past the home page of the website) (Scassa and Deturbide, 2012). TOS frequently address rules regarding the use of content on the site. For example, Airbnb's TOS provide users with only a limited right to access and view the site's content for personal and non-commercial purposes (TOS, Art. 4). Also prohibited is the circumvention of any TPMs (TOS, Art. 14.1). Copying or adapting content is not permitted (TOS, Art. 5.3). Data scraping or any other form of automated data extraction is also specifically prohibited (TOS, Art. 14.1).

Because TOS are contracts, they have an important limitation. Generally, contracts only bind the parties to the agreement. Thus, the fact that data have been scraped in breach of the contract between the platform and the data scraper does not affect a third party who uses the scraped data. However, where there are also copyright claims in the scraped content, a third party may be restricted in their use of the content by these property-based claims. And, as was discussed above in the section on rights to use, the existence of a contractual obligation to not scrape data might be a factor in assessing whether data scraping that breaches intellectual property rights is fair use/fair dealing. The Court of Justice of the European Union has recently held that contractual TOS that prohibit scraping may provide a basis for finding liability for breach of contract (*Ryanair Ltd* v. *PR Aviation BV*, 2015), and a similar result was reached in a Canadian case (*Century 21 Canada Limited Partnership* v. *Rogers Communications Inc*., 2011).

Where technological barriers are in place, the circumvention of these barriers may lead to different legal consequences. Anti-circumvention provisions, now found in most copyright statutes, provide additional recourse to a plaintiff where a defendant circumvents TPMs to gain access to copyright-protected content (Craig, 2010). A TPM may be as simple as a username and password for a site (Puerta, 2016). In the case of most publicly accessible data, such technological restrictions will not exist. There has been some discussion regarding

whether ignoring a robots.txt protocol – designed to communicate to web crawlers and robots that they are not permitted on the site – constitutes circumvention (Lundblad, 2007), although this would seem to be a relatively weak argument (Jasiewicz, 2012).

In the USA, the *Computer Fraud and Abuse Act* has been invoked in a number of data-scraping cases. This statute makes it a criminal offense to make unauthorized use of computers. While the statute was originally intended to provide recourse against hackers, it has been invoked in cases involving data scrapers. There has been some debate as to whether data scraping is actually captured by this statute, which seems more tailored to address breaches of security measures in order to obtain confidential data rather than to scraping publicly accessible data from websites (Din, 2015; Hirschey, 2014). Some data-scraping prosecutions under the CFAA have succeeded, while others have failed, creating considerable uncertainty (Din, 2015). Din (2015) argues that the CFAA should be confined to those data-scraping cases which involve the circumvention of some kind of technological barrier such as password protection or encryption, but should not apply to the scraping of publicly accessible information. This view seems to be reflected in the recent decision of the US District Court for the District of Columbia in *Sandvig* v. *Sessions* (2018). The Court found that the plaintiffs had standing to sue and to argue that the application of the CFAA to data scraping for research purposes would violate the Free Speech and the Free Press clauses of the US Constitution.

It is interesting to note that in the case of publicly accessible information, many of the platform companies' recourses, described above, have weaknesses, and these are often significant. Copyright claims in compilations of data may be of questionable scope or strength, and trespass to chattels claims will be weak where there is no particular impact on the server. Anti-circumvention claims will depend upon the existence of effective TPMs as well as copyright-protected content. The CFAA, applicable only in the USA in any event, is also controversial in its application in these contexts. Nevertheless, the fact that such recourses exist can be a strong deterrent, particularly where there is disparity in economic power between the host platform and the data scraper. In such contexts, the recipient of a cease-and-desist letter may have little option but to cease-and-desist since any attempt to resist the asserted claims will require considerable financial resources.

As a result, the legal uncertainty as to the status of publicly accessible data and the scope of users' rights could significantly inhibit the actions of many users in the face of even the slightest pushback by the platform company. And, as noted earlier, if under-resourced users are unable to litigate in support of these interests, any case law in this area will continue to evolve between commercial actors without adequate attention to users' rights.

Privacy

Much publicly accessible platform data are also personal information. On a platform such as Airbnb, some user-contributed data may be personal information. Certainly some personal information about hosts is visible on the site, and guests may share personal information in the form of reviews of places where they have stayed. Although not all such information is accompanied by the full name of an individual, all that is required for data protection laws to apply is that individuals be identifiable. The scraping of such data raises important and challenging privacy issues. In some cases, publicly accessible platform data might be scraped specifically for commercial purposes relating to profiling and targeting of individuals (Canales, 2018). Thus, to the extent that publicly accessible platform data include personal information, data protection laws may impose additional restrictions on the collection, use and disclosure of these data.

Different jurisdictions take different approaches to personal information that is made publicly available on websites. In the USA, such data are generally considered fair game, since they have been made public by the data subject and therefore do not attract a

reasonable expectation of privacy, although some have challenged this view (Scott, 2017). The Airbnb Privacy Policy makes it clear that content users contribute to public portions of the site is "visible to the general public" (Art. 3.3). Nevertheless, although the information may be visible to the general public, Airbnb's TOS prohibit the downloading, scraping or other extraction of data from its site. In theory, therefore, although users agree to share data publicly for the purposes of facilitating the rental of accommodations, they do not consent to this information being downloaded or scraped and used for other purposes. One question is therefore whether a platform could assert the privacy rights of its users as a basis for legal action against data scrapers. This argument was raised in the litigation between hiQ and LinkedIn. LinkedIn – a site that also makes a great deal of user personal information available for public viewing – asserted its privacy commitments to its users as a basis for attacking hiQ's scraping of its website contents. In deciding a preliminary proceeding brought by hiQ, a California court expressed doubt about the extent of LinkedIn users' expectation of privacy in the publicly accessible content, and also noted that the platform's own actions with respect to its users' privacy did not appear to have been particularly "zealous" (*hiQ Labs, Inc.* v. *LinkedIn Corporation*, 2017). Although it is not clear that the same approach would be taken in jurisdictions with different data protection laws, the court's approach highlights the difficulties with raising privacy issues with respect to personal information available over a publicly accessible platform.

In Canada, private sector data protection laws would apply to the collection of personal information by a company for commercial purposes, even if that information were publicly available platform data. However, with an eye to impending reforms of the federal statute, there has been some discussion of whether such data should be excluded from consent requirements (see, e.g. Canada, Standing Committee on Access to Information, Privacy and Ethics, 2018). In the EU, publicly accessible platform data would be subject to the General Data Protection Regulation, although the public nature of user-contributed personal information might have an impact on the degree of protection available.

Conclusion

As the Airbnb example demonstrates, a diverse range of users (including researchers, journalists, competing and non-competing businesses) make use of publicly accessible platform data for multiple purposes, many of which serve a broader public interest. These uses are subject to challenge by the platform companies that assert legal rights of ownership and control. The existing statute and case law that buttress claims of ownership/control may also provide a framework of sorts for user rights, yet this framework is not well adapted to our evolving data society generally or to platform data ecosystems in particular. The economic and power imbalances that always impact the litigation process can be exacerbated where existing laws are interpreted and applied to rapidly evolving contexts. There is a considerable risk that such power imbalances can mean that user perspectives and the public interest will not be well represented in the evolving litigation, if they are represented at all. Thus, the broader concern is how to ensure meaningful access to online data in the public interest when it is in private sector hands.

Attempts to control publicly accessible platform data must be seen in the context of the complex ecosystems that can arise around them, as illustrated by the diverse user-base for Airbnb data. An ecosystem approach is particularly useful to address the reality that, for the time being, important issues around the legitimacy of data-scraping activities are likely to be decided by litigation between commercial competitors. The high cost of litigation means that non-commercial users such as researchers and non-profit organizations, as well as small start-up companies, are unlikely to pursue issues in court. An ecosystem approach keeps the diversity of users and uses of publicly accessible

data at the forefront and can help shape more nuanced approaches to the issues. This is particularly important in a novel and rapidly evolving data context.

Different legislative solutions are available. One might be to provide clarification of the scope of protection available to compilations of data, including publicly accessible platform data. Data scraping could be addressed in a new copyright exception similar to what is being considered in the EU for text and data-mining activities. However, there are risks in taking such an approach since new exceptions have the potential of limiting reuse in the public interest simply by casting their scope too narrowly. In a country such as the USA, where fair use is a broad and flexible exception, a new exception might be less desirable than in a country such as Canada, where fair dealing is limited to specific contexts and where the public interest has often fared poorly in the hands of lower courts. Regardless, any new legislative measures must take into account the complex ecosystems that emerge around publicly accessible data, including the broad range of potential users and uses of the data. Legislative amendments to prevent bulk contracting out of fair use/fair dealing rights could also be important in this context, particularly since scraping may be prohibited outright by TOS, as is the case with Airbnb. Furthermore, TPMs in copyright law should not be interpreted so broadly as to encompass tools such as the robots.txt protocol where data are otherwise publicly available. Laws such as the CFAA in the USA should be interpreted narrowly so as to not capture scraping of publicly accessible data.

It is worth noting that the combination of barriers erected to data scrapers and the laws that reinforce them can raise ethical issues. The issue of the ethics of data scraping has already arisen in some contexts such as journalism (Shiab, 2015). It has also arisen in relation to research ethics (Fiesler, 2017; Bruckman, 2016), although some research ethics concerns have focused on the privacy implications of the use of such data (Zimmer, 2010), and not on the ethical implications of researchers breaching contractual terms of use in order to access data, or even engaging in activities that might be categorized as tortious (trespass to chattels). The evolution of the law in this area could have a significant impact on how ethical issues are addressed, and this, in turn, could lead to further restrictions on the ability of institutional researchers to make use of publicly accessible data. Even the legal uncertainties on their own can be stifling, particularly as threats of legal action combined with disparity in economic power can suppress uses/activity unless there are clear rights to access or use.

There is no doubt that these legal uncertainties will need to be resolved. This can be done through clear public policy direction from governments. The discussions in the EU over the creation of a new text and data-mining exception show that concrete action is possible to address the impact of new technologies on the copyright balance. However, such action tends to be driven by commercial interests. User interests rarely prompt swift legislative responses. Failing concrete action, the law will evolve on an incremental basis, driven by the litigation strategies of major corporate players. Within this context, an ecosystem approach to publicly accessible data becomes essential to set the context in which competing claims should be assessed. The ecosystem approach has the advantage of broadening the analysis beyond the specific claims of parties to litigation and considering instead at the diverse ways in which publicly accessible data are used and the broader public interests that may be served by such uses.

References

Airbnb (n.d.), "About Us", available at: www.airbnb.ca/about/about-us (accessed February 12, 2018).

Associated Press v. *Meltwater U.S. Holdings, Inc.* (2013), "931 F Supp (2d) 537", (SDNY 2013).

Bruckman, A. (2016), "Do researcher's need to abide by terms of service (TOS)? An answer", The Next Bison: Social Computing and Culture, February 26, available at: https://nextbison.wordpress.com/2016/02/26/tos/ (accessed February 12, 2018).

Canada, Standing Committee on Access to Information, Privacy and Ethics (2018), "Towards privacy by design: review of the personal information protection and electronic documents act", 42nd Parliament, 1st Session, February 12, available at: www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-12/ (accessed July 5, 2018).

Canales, K. (2018), "What is the Facebook public profile information that could have been scraped?", Business Insider, April 4, available at: www.businessinsider.com/what-is-facebook-public-profile-information-that-could-be-scraped-2018-4 (accessed July 5, 2018).

Century 21 Canada Limited Partnership v. *Rogers Communications Inc.* (2011), "BCSC 1196 (CanLII)", available at: http://canlii.ca/t/fn00h (accessed February 12, 2018).

City of Toronto (2016), "Developing an approach to regulating short-term rentals", October 11, City of Toronto, Toronto, ON, available at: www.toronto.ca/legdocs/mmis/2016/ex/bgrd/backgroundfile-97235.pdf (accessed February 12, 2018).

Clampet, J. (2014), "Airbnb in NYC: the real numbers behind the sharing story", Skift, February 13, available at: https://skift.com/2014/02/13/airbnb-in-nyc-the-real-numbers-behind-the-sharing-story/ (accessed February 12, 2018).

Computer Fraud and Abuse Act (1984), "18 U.S.C. § 1030".

Conger, K. (2016), "LinkedIn sues anonymous data scrapers", TechCrunch, August 15, available at: https://techcrunch.com/2016/08/15/linkedin-sues-scrapers/ (accessed July 5, 2018).

Cox, K.L. (2015), "Issue brief: text and data mining and fair use in the United States", Association of Research Libraries, June 5, available at: www.arl.org/storage/documents/TDM-5JUNE2015.pdf (accessed July 5, 2018).

Cox, M. (2017), "The face of Airbnb: New York City", March 1, available at: http://insideairbnb.com/face-of-airbnb-nyc/ (accessed February 12, 2018).

Craig, C. (2010), "Locking out lawful users: fair dealing and anti-circumvention in bill C-32", in Geist, M. (Ed.), *From Radical Extremism to Balanced Copyright: Canadian Copyright and the Digital Agenda*, Irwin Law, Toronto, pp. 177-203.

Cutler, K. (2015), "Airbnb, proposition F and the shared hypocrisy of bay area housing", TechCrunch, November 3, available at: https://techcrunch.com/2015/11/03/prop-f/ (accessed February 12, 2018).

Di Valentino, L. (2014), "Conflict between contract law and copyright law in Canada: do licence agreements Trump users' rights?", FIMS Working Papers, Paper No. 1, London, ON, available at: http://ir.lib.uwo.ca/fimswp/1 (accessed February 12, 2018).

Din, M.F. (2015), "Breaching and entering: when data scraping should be a federal computer hacking crime", *Brooklyn Law Rev*, Vol. 81 No. 1, pp. 405-440.

Edelman, B., Luca, M. and Svirsky, D. (2017), "Racial discrimination in the sharing economy: evidence from a field experiment", *American Economic Journal*, Vol. 9 No. 2, pp. 1-22, available at: www.aeaweb.org/articles?id=10.1257/app.20160213 (accessed February 12, 2018).

Elkin-Koren, N. (1997), "Copyright policy and the limits of freedom of contract", *Berkeley Technology Law Journal*, Vol. 12 No. 1, pp. 93-113, available at: https://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=1144&context=btlj (accessed February 12, 2018).

Elmahjub, E. and Suzor, N. (2017), "Fair use and fairness in copyright: a distributive justice perspective on users' rights", *Monash University Law Review*, Vol. 43 No. 1, pp. 274-298.

Fiesler, C. (2017), "Law & ethics of scraping: what *hiQ* v. *LinkedIn* could mean for researchers violating TOS", Medium, August 15, available at: https://medium.com/@cfiesler/law-ethics-of-scraping-what-hiq-v-linkedin-could-mean-for-researchers-violating-tos-787bd3322540 (accessed February 12, 2018).

Geiger, C., Frosio, G. and Bulayenko, O. (2018), "The exception for text and data mining (TDM) in the proposed directive on copyright in the digital single market – legal aspects", Centre for International Intellectual Property Studies, Université de Strasbourg, Strasbourg, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3160586 (accessed July 6, 2018).

Harrison, T.M., Pardo, T.A. and Cook, M. (2012), "Creating open government ecosystems: a research and development agenda", *Future Internet*, Vol. 4 No. 4, pp. 900-928, available at: www.ctg.albany.edu/publications/journals/og_ecosystems_2012/og_ecosystems_2012.pdf (accessed February 12, 2018).

Hiltzik, M. (2015), "No surprise: that Airbnb study of rentals in L.A. isn't what it seems", *Los Angeles Times*, September 30, available at: www.latimes.com/business/hiltzik/la-fi-mh-airbnb-study-of-rentals-20150930-column.html (accessed February 12, 2018).

hiQ Labs, Inc. v. *LinkedIn Corporation* (2017), "273 F. Supp. 3d 1099 – Dist. Court", ND CA.

Hirschey, J.K. (2014), "Symbiotic relationships: pragmatic acceptance of data scraping", *Berkeley Technology Law Journal*, Vol. 29 No. 1, pp. 897-928.

Host Compliance LLC (2016), "City of Vancouver: short-term rental market overview", available at: http://vancouver.ca/files/cov/overview-vancouver-short-term-rental-market.pdf (accessed February 12, 2018).

Hugenholtz, P.B. (2017), "Data property: unwelcome guest in the house of IP", paper presented at Trading Data in the Digital Economy: Legal Concepts and Tools, Münster, available at: https://pure.uva.nl/ws/files/16856245/Data_property_Muenster.pdf (accessed February 12, 2018).

Information Commissioner's Office (2018), "Investigation into the use of data analytics in political campaigns", available at: https://ico.org.uk/media/action-weve-taken/2259371/investigation-into-data-analytics-for-political-purposes-update.pdf (accessed December 10, 2018).

Jamasi, Z. and Hennessy, T. (2016), *Nobody's Business: Airbnb in Toronto*, Canadian Centre for Policy Alternatives, Toronto.

Jasiewicz, M.I. (2012), "Copyright protection in an opt-out world: implied license doctrine and news aggregators", *Yale Law Journal*, Vol. 122 No. 3, pp. 837-850.

Leaffer, M. (2007), "Database protection in the United States is alive and well: comments on Davison", *Case Western Reserve Law Review*, Vol. 57 No. 4, pp. 855-862.

Lundblad, N. (2007), "e-Exclusion and bot rights: legal aspects of the robots exclusion standard for public agencies and other public sector bodies with Swedish examples", *First Monday*, Vol. 12 No. 8, available at: http://firstmonday.org/ojs/index.php/fm/article/view/1974/1849 (accessed February 12, 2018).

Majoribanks, I. (2016), "Airbnb in Vancouver and its impacts on affordable housing", Affordable Vancouver, June 1, available at: https://affordablevancouver.files.wordpress.com/2016/07/airbnb-vancouver-report-iain-marjoribanks.pdf (accessed February 12, 2018).

Nardi, B. and O'Day, V.L. (1999), *Information Ecologies: Using Technology with Heart*, MIT Press, Cambridge, MA.

Newell, B.C. (2011), "Out with the old and in with the new: converging standards of originality for database protection", *University of San Francisco Intellectual Property Law Bulletin*, Vol. 15, pp. 111-126.

Newspaper Licensing Agency Ltd and Ors v. *Meltwater Holding BV and Ors* (2011), "EWCA Civ 890".

Office of the Attorney General of the State of New York (2014), "Airbnb in the city", available at: https://ag.ny.gov/pdfs/AIRBNB%20REPORT.pdf (accessed February 12, 2018).

Puerta, M.P. (2016), "Public policies for education in Latin America and the difficulties imposed by international obligations for technological protection measures", *American University International Law Review*, Vol. 32 No. 1, pp. 165-210.

Reyes, E.A. (2015), "New soldiers in Airbnb battle: PR and politics", *Los Angeles Times*, April 4, available at: www.latimes.com/local/politics/la-me-adv-airbnb-politics-20150405-story.html#page=1 (accessed February 12, 2018).

Ryanair Ltd v. *PR Aviation BV* (2015), "Case C-30/14", January 15.

Said, C. (2015), "The Airbnb effect", *San Francisco Chronicle*, July 12, available at: www.sfchronicle.com/airbnb-impact-san-francisco-2015/#1; www.sfchronicle.com (accessed February 12, 2018).

Sandvig v. *Sessions* (2018), "Civil action No. 16-1368", available at: www.aclu.org/legal-document/sandvig-v-sessions-opinion (accessed July 5, 2018).

Sawatzky, K. (2015), "Airbnb listings in Vancouver: how many? What type? Where?", Short-Term Consequences, Vancouver, BC, June 20, available at: http://shorttermconsequences.wordpress.com/2015/06/20/airbnb-listings-in-vancouver-how-many-what-type-where (accessed February 12, 2018).

Sawatzky, K. (2016), "Short-term consequences: investigating the extent, nature and rental housing implications of Airbnb rentals listings in Vancouver", master's MUrb thesis, Simon Fraser University Faculty of Arts and Applied Sciences, available at: http://summit.sfu.ca/item/16841#310 (accessed February 12, 2018).

Scassa, T. (2017), "Sharing data in the platform economy: a public interest argument for access to platform data", *University of British Columbia Law Review*, Vol. 50 No. 4, pp. 1017-1071.

Scassa, T. and Deturbide, M. (2012), *Electronic Commerce and Internet Law in Canada*, 2nd ed., CCH/Wolters Kluwer, Toronto.

Scott, J.D. (2017), "Social media and government surveillance: the case for better privacy protections for our newest public space", *Journal of Business and Technology Law*, Vol. 12 No. 2, pp. 151-164.

Shiab, N. (2015), "On the ethics of web scraping and data journalism", J-Source, June 22, available at: www.j-source.ca/article/ethics-web-scraping-and-data-journalism (accessed February 12, 2018).

Snell, J. and Care, D. (2013), "Use of online data in the Big Data era: legal issues raised by the use of web crawling and scraping tools for analytics purposes", Bloomberg Law, August 28, available at: www.bna.com/legal-issues-raised-by-the-use-of-web-crawling-and-scraping-tools-for-analytics-purposes (accessed February 12, 2018).

Sobel, B. (2017), "Artificial intelligence's fair use crisis", *Columbia Journal of Law and the Arts*, Vol. 41 No. 1, pp. 45-98.

Stulberg, A. (2016), "Airbnb probably isn't driving rents up much, at least not yet", FiveThirtyEight, August 24, available at: https://fivethirtyeight.com/features/airbnb-probably-isnt-driving-rents-up-much-at-least-not-yet/ (accessed February 12, 2018).

Tamaroff, D. (2011), "Bottling the free flow of information: a comparative analysis of U.S. and E.U. database protection", *Wake Forest Journal Business and Intellectual Property Law*, Vol. 12, pp. iii-24.

Trader Corporation v. *CarGurus, Inc.* (2017), "ONSC 1841".

Vancouver City Council (2016), "Motion: action to strengthen regulation of short-term rentals", April 5, available at: http://council.vancouver.ca/20160406/documents/pspc10.pdf (accessed February 12, 2018).

Wang, D., Xi, S. and Gilheany, J. (2015), "The model minority? Not on airbnb.com: a hedonic pricing model to quantify racial bias against Asian Americans", Technology Science, September 1, available at: https://techscience.org/a/2015090104 (accessed February 12, 2018).

Warner, R. (2002), "Border disputes: trespass to chattels on the internet", *Villanova Law Review*, Vol. 47 No. 1, pp. 117-159.

Zimmer, M. (2010), "'But the data is already public': on the ethics of research in Facebook", *Ethics and Information Technology*, Vol. 12 No. 4, pp. 313-325, available at: https://link.springer.com/article/10.1007/s10676-010-9227-5 (accessed February 12, 2018).

Zuiderwijk, A., Janssen, M. and Davis, C. (2014), "Innovation with open data: essential elements of open data ecosystems", *Information Polity*, Vol. 19 No. 1, pp. 17-33, available at: https://content.iospress.com/articles/information-polity/ip000329 (accessed February 12, 2018).

Further reading

Airbnb (2017a), "Airbnb privacy policy", June 19, available at: www.airbnb.ca/terms/privacy_policy (accessed February 12, 2018).

Airbnb (2017b), "Terms of service", June 19, available at: www.airbnb.ca/terms/ (accessed February 12, 2018).

Brauneis, R.F. (Ed.) (2009), *Intellectual Property Protection of Fact-based Works: Copyright and its Alternatives*, Edward Elgar, Cheltenham.

City and County of San Francisco Board of Supervisors Budget and Legislative Analyst (2015), "Policy analysis report", May 13, available at: http://sfbos.org/sites/default/files/FileCenter/Documents/52601-BLA.ShortTermRentals.051315.pdf (accessed February 12, 2018).

Green, M.S. (2009), "Two fallacies about copyrighting factual compilations", in Brauneis, R.F. (Ed.), *Intellectual Property Protection of Fact-based Works: Copyright and its Alternatives*, Edward Elgar, Cheltenham, pp. 109-132.

Hughes, J. (2007), "Created facts and the flawed ontology of copyright law", *Notre Dame Law Review*, Vol. 83 No. 1, pp. 43-108.

San Francisco Planning Department (2015), "Executive summary: administrative code text change", April 16, available at: http://commissions.sfplanning.org/cpcpackets/2014-001033PCA.pdf (accessed February 12, 2018).

Scassa, T. (2006), "Original facts: skill, judgment and the public domain", *McGill Law Journal*, Vol. 51 No. 2, pp. 253-278.

About the author

Dr Teresa Scassa is Canada Research Chair in Information Law and Policy at the University of Ottawa. She is a member of the GEOTHINK research partnership, and has written widely in the areas of intellectual property law, law and technology and privacy. Dr Teresa Scassa is also Senior Fellow with CIGI's International Law Research Program. She is a founding member of the University of Ottawa's Centre for Law, Technology and Society, is cross-appointed to the School of Information Studies at the University of Ottawa and is a member of the Geomatics and Cartographic Research Centre at Carleton University. Dr Teresa Scassa can be contacted at: tscassa@uottawa.ca